

## SEARCH REQUEST FORM

Examiner # (Mandatory): \_\_\_\_\_ Requester's Full Name: \_\_\_\_\_

Art Unit \_\_\_\_\_ Location (Bldg/Room#): 9E07 Phone (circle 305 306 308) \_\_\_\_\_

Serial Number: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Date: \_\_\_\_\_

Keywords (include any known synonyms registry numbers, explanation of initialisms): \_\_\_\_\_

## Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

## STAFF USE ONLY

Searcher: <u>D. Schreiber</u>	Type of Search	Vendors (include cost where applicable)
Searcher Phone #: <u>308-4292</u>	<u>12</u> N.A. Sequence	_____ STN
Searcher Location: <u>CM 12 E18</u>	<u>1</u> A.A. Sequence	_____ Questel/Orbit
Date Picked Up: <u>7/13</u>	_____ Structure (#)	_____ Lexis/Nexis
Date Completed: <u>7/16</u>	_____ Bibliographic	_____ WWW/Internet
Classical Prep Time: <u>8</u>	_____ Litigation1	<u>✓</u> In-house sequence systems (list) <u>MP Search</u>
Terminal Time: <u>11</u>	_____ Fulltext	_____ Dialog
Number of Databases: <u>7</u>	_____ Procurement	_____ Dr. Link
	_____ Other	_____ Westlaw
		_____ Other (specify)

**THIS PAGE BLANK (USPTO)**

## IMPORTANT INFORMATION ABOUT YOUR SEQUENCE SEARCH:

### Compugen Sequence searching hardware and software explained:

This is the new sequence searching system that is currently being phased into as a replacement for the Maspar/Mpsrch platform. This system has been tested by both searchers and examiners, and has shown equivalent results to the Maspar system for the same databases. The results output format for all Compugen printed results are essentially the same except for translations.

### Translation searching on Compugen explained:

The Compugen system utilizes Framesearch software for translations of proteins to nucleotides, and nucleotides to proteins. Some examiners have found these to be superior to the backtranslate software on Maspars.

FrameSearch searches a group of protein sequences for similarity to one or more nucleotide query sequences, or searches a group of nucleotide sequences for similarity to one or more protein query sequences. For each sequence comparison, the program finds an optimal alignment between the protein sequence and the corresponding codons on each the nucleotide sequence. Optimal alignments may include reading frame shifts. Please see any of the professional searching staff if you need assistance with this format.

### File extensions for Compugen results transferred to floppy disks.

Compugen system search results will be delivered in one of two possible formats:

1. Standard concatenated files with .flp extension.
2. Compressed .zip files which decompressed yield two files as described below:

US08123456.cmr - Contains all commercial databases, may include Issued  
US08123456.pen - Contains pending file results only

## VERY IMPORTANT NOTE ABOUT PENDING FILE SEARCHES.

If your search contains file names with the following bolded extensions:

US08123456.rap      US08123456.rnp

Do not leave this search in the case, during prosecution, or after the case issues, since it contains pending data which is confidential.

### **QUESTIONS? Contact any of the following:**

Dilip Pandya, Chief, Information Branch, 308-4268

#### Professional searching staff:

John Dantzman (308-4488); Jan Delaval (308-4498); Mary Hale (308-4258); Barb O'Bryen (308-4291); David Schreiber (308-4292); Paula Sheppard (308-4499); Mark Spencer (308-4266); Beverly Shears (308-4994); Alex Wacławiw (308-4491).

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 12:50:38 ; Search time 822.9 Seconds  
(without alignments)  
4334.446 Million cell updates/sec

Title: US-09-156-580-1

Perfect score: 997

Sequence: 1 cccagtgccatttttctct.....aaagtaataataaaaaaa 997

Scoring table: IDENTITY\_NUC

Searched: 808301 seqs, 1798773984 residues

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl1.\*
- 9: gb\_pl2.\*
- 10: gb\_prl.\*
- 11: gb\_pr2.\*
- 12: gb\_pr3.\*
- 13: gb\_ro.\*
- 14: gb\_sc.\*
- 15: gb\_sy.\*
- 16: gb\_un.\*
- 17: gb\_vi.\*
- 18: gb\_htg.\*
- 19: em\_ba.\*
- 20: em\_fun.\*
- 21: em\_hum1.\*
- 22: em\_hum2.\*
- 23: em\_in.\*
- 24: em\_om.\*
- 25: em\_or.\*
- 26: em\_ov.\*
- 27: em\_pat.\*
- 28: em\_ph.\*
- 29: em\_pl.\*
- 30: em\_ro.\*
- 31: em\_sy.\*
- 32: em\_un.\*
- 33: em\_vi.\*
- 34: em\_htg.\*
- 35: em\_sts.\*
- 36: gb\_bal.\*
- 37: gb\_ba2.\*
- 38: gb\_pl1.\*
- 39: gb\_pl2.\*
- 40: gb\_prl.\*
- 41: gb\_pr2.\*
- 42: gb\_pr3.\*
- 43: gb\_sts.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

ALIGNMENTS

RESULT 1

ATU38946

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

1	89	8.9	1071	9	ATU38946	U38946 Arabidopsis
2	89	8.9	1071	39	ATU38946	U38946 Arabidopsis
3	76.2	7.6	99687	18	AC005956	AC005956 *** SEQUE
4	70.4	7.1	93845	9	ATAC004684	AC004684 Arabidops
5	70.4	7.1	93845	39	ATAC004684	AC004684 Arabidops
6	49	4.9	5931	3	DDU32174	U32174 Dictyostell
7	47.2	4.7	38829	2	AE000787	AE000787 Borrelia
8	47.2	4.7	38829	37	AE000787	AE000787 Borrelia
9	46.8	4.7	1200	8	MISCV	V00705 Yeast mitoc
10	46.8	4.7	1200	8	YSCMTVAR1	J01525 Yeast (S.ce
11	46.8	4.7	1200	38	MISCV	V00705 Yeast mitoc
12	46.8	4.7	1200	38	YSCMTVAR1	J01525 Yeast (S.ce
13	46.2	4.6	1083	8	PETZFD82	D26084 Petunia mRN
14	46.2	4.6	1083	38	PETZFD82	D26084 Petunia mRN
15	45.6	4.6	58407	2	MI12CG	L77118 Methanococc
16	45.6	4.6	58407	37	MI12CG	L77118 Methanococc
17	44.8	4.5	251364	13	MMAE000664	AE000664 Mus muscu
18	44.6	4.5	3041	8	PETZFD81	D26083 Petunia zin
19	44.6	4.5	3041	38	PETZFD81	D26083 Petunia zin
20	44.4	4.5	1453	3	DD1039PPR1	AF018638 Dictyoste
21	44.2	4.4	12688	3	AF079445	AF079445 Dictyoste
22	44	4.4	28768	3	PFSC03020	AL008972 Plasmodiu
23	43.8	4.4	7467	3	PFSC04099	AL010282 Plasmodiu
24	43.8	4.4	5100	8	YSCPH081	D13228 Yeast PHOE-
25	43.8	4.4	5100	38	YSCPH081	D13228 Yeast PHOE-
26	43.4	4.4	11442	3	CEH06A10	D13228 Yeast PHOE-
27	43.2	4.3	4374	3	AF045453	AF045453 Dictyoste
28	42.2	4.2	553	3	AF010544	AF010544 Plasmodiu
29	42.2	4.2	1139	3	AF033931	AF033931 Bactrocer
30	42.2	4.2	6265	3	PFSC03080	AL010153 Plasmodiu
31	42.2	4.2	3478	8	DDIADCYG	M87278 Dictyostelli
32	42.2	4.2	3478	38	DDIADCYG	M87278 Dictyostelli
33	41.8	4.2	3700	3	DGP860G	X66483 D.discoiden
34	41.8	4.2	235	3	PFPPF297	X53020 P.falciparu
35	41.6	4.2	38383	3	CEF35H8	Z36752 Caenorhabdi
36	41.4	4.2	6115	3	DDU20432	D20432 Dictyostelli
37	41.4	4.2	15421	3	PFSCMP1A	X95275 P.falciparu
38	41.4	4.2	3075	3	XPFRNAPOL	X75544 P.falciparu
39	41.4	4.2	165	43	G37866	G37866 emaa2 Plasm
40	41.2	4.1	2107	8	DDIADCVA03	L05498 Dictyostelli
41	41.2	4.1	114827	11	AC004413	AC004413 Homo sapi
42	41.2	4.1	2107	38	DDIADCVA03	L05498 Dictyostelli
43	41.2	4.1	114827	41	AC004413	AC004413 Homo sapi
44	41.2	4.1	221	43	G37927	G37927 c3m61 Plasm
45	41	4.1	2696	8	DDIPKNGPK	M38703 Dictyostelli

ATU38946 1071 bp DNA PLN 29-NOV-1995  
Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.

U38946

g1079668

thale cress.

Arabidopsis thaliana

Eukaryotae; mitochondrial eukaryotes; Viridiplantae;

Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;

Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.

1 (bases 1 to 1071)

Sakai, H., Medrano, L.J. and Meyerowitz, E.M.

Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries

Nature 378 (6553), 199-203 (1995)

96069740

2 (bases 1 to 1071)

Sakai, H., Medrano, L.J. and Meyerowitz, E.M.

Direct Submission

Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech,

Pasadena, CA 91125, USA

Location/Qualifiers





repeat_region	QSLTPCSVENHYESEKYAKDILDEPKRIATKLLTSHVSGTGCHKKHVKSEGSNKRKRNDLHRLFPNGLPDGTELAYVYKTOISPSQFEAHAGMAARQPYRHIPISISGLSLHDIAM		repeat_region	TGGGFL" 11697. .11757		
	SLANGHVIITGDSDMCSCGGDGLLACGCPQAFHTACLKQSPMEGTWYCSCSND			/rpt_family=" (CAT)n"		
	GPISKKATTPDPSGNARPIVIRLSRVKAPESDIGCVFCRSHDFSIGKFDORTVIL			13716. .13952		
	CQCEKEYHVGCGLRENGCDLKEIPOEKWFCSCNCRIHTAVONSVCSPQTLPTPLL			/note="exon predicted by xgrail, quality excellent_shadowexon"		
misc_feature	DMICRDKREKIFTDGDVREHILSGKRYPEHLPLLSRAAVIFRECDDPIVAKSGR		misc_feature	complement(14172. .14309)		
	DLIPVMYGRNISGGFGGMVCLLVNSLVVSAALLRIFGQGEVALP1VATVSRYQG			/note="exon predicted by xgrail, quality marginal"		
	RIFYOGLYACVENLSSLVNENLVLPAAEEASITWKKEGFTKMSDQQLQEQYKEVOL			join(14707. .14911,15566. .15765,15858. .15964,16057. .16189,16311. .16602,16755. .17182)		
	TIFKGTSMLEKKVPKATTGLSESTLI"			/gene="F13M22.6"		
repeat_region	/rpt_family=" (GGAA)n"		gene	14707. .17294		
misc_feature	6023. .6202		/note="identical to GB:U40269"			
repeat_region	marginal_shadowexon"		repeat_region	/gene="F13M22.6"		
	6179. .6236			join(15634. .15765,15858. .15964,16057. .16189,16311. .16602,16755. .17182)		
	/rpt_family="POLY_A"			/gene="F13M22.6"		
	join(<6647. .6702,7161. .7304,7322. .>7499)			/codon_start=1		
gene	/gene="F13M22.3"		/product="origin recognition complex protein"			
CDS	<6647. .>7499		CDS	/db_xref="PID:g3236239"		
	/note="predicted by genscan"			/translation="MEDIENTEEDYGFSRNYFLAKELGGASKRAHKLSDIHIIVDEQ		
	join(6647. .6702,7161. .7304,7322. .7499)			ELRETPASTIEMKHSKEISELMSDYKTMYSKWVEELRCGFLLMYCGSKKALVEDFAS		
	/gene="F13M22.3"			ASLTIDYSVYVINGILPSVNLQVLLAELSELCKCRKSSGSLSKGQETFPSPRSM		
mRNA	/note="hypothetical protein"		mRNA	DILSLFHGPQSGKDCFCVYVHNIDGPALRDPESQQTARLSSSHRLVASIDHVN		
	/codon_start=1			APLMDKKVHKQENLWHVPTFAPYINVEGVFPLVLAQGSTAQTAKTAIVLQSLT		
	/db_xref="PID:g3236236"			PNGQVFKILAEYQLSHDPEDGMPDLDLYSASRERFFVSQVTLNSHLTEFKDHELK		
	/translation="MTWKVKREENKSFQFLFKFARIPHFVYGATWTLFLTITVAII			TKRNSDGOCLNIPLTSDAIRQLLDLQ"		
gene	SLAPEAFVSAIFPSSEVSFRRHIGSAGSLGLVFTGESFEVEDGPRCAAGFCN		gene	complement(<17532. .>19016)		
	CGSFRRRKRDYGLGRRSPLI"			complement(<17532. .>19016)		
	complement(join(<7761. .7931,8019. .8065,8154. .8238,8323. .8401,8494. .8647,8824. .8903,9001. .9232,9769. .>9886))			/gene="F13M22.7"		
	/gene="F13M22.4"			/note="unknown protein"		
gene	complement(<7761. .>9886)		CDS	/codon_start=1		
	/gene="F13M22.4"			/db_xref="PID:g3236240"		
	complement(join(7761. .7931,8019. .8065,8154. .8238,8323. .8401,8494. .8647,8824. .8903,9001. .9232,9769. .>9886))			/translation="MENHHPSTLLSMDSASSHEELDEMNNRSLSGPPDINLPL		
	/gene="F13M22.4"			SAERPPPPPNWNLDACDILDVGLGSOAYETENYMSVPKVRKCAKRVDSIWGANFFF		
repeat_region	/product="putative ribitol dehydrogenase"		repeat_region	/note="exon predicted by xgrail, quality excellent_shadowexon"		
	/db_xref="PID:g3236237"			join(<20811. .21507,21667. .21805,22028. .22434,22931. .23006,23118. .23233,23331. .>23788)		
	/translation="MGIVGMTCKKSGFGSASTAEDVTOAIDASHLTAITIGTSG			/gene="F13M22.8"		
	IGLEAARVLAMRGHVIIAARPKANESKEMILQNPARNVDYIQLDINSSLSKVSF			<20811. .>23788		
repeat_region	VDQFLANVPLNLIINAGNCPFKLTEDGESOFATNHGIELTNLLDKMSTA		misc_feature	/gene="F13M22.8"		
	RSQVGQIVNLSIAHTYTYSEGKFOGINDPAGYSERRAYCQSKLSNLLHSNLSR			join(20811. .21507,21667. .21805,22028. .22434,22931. .23006,23118. .23233,23331. .23788)		
	RLOEGVNTINSVHPGLVTNLFRYSGFSMKVFRAMTELFWKNIPOGAATCYVALH			/codon_start=1		
	PLOGVGTGYGDCNIVAPSKFATNNSLADKLWDSFVLIDSIS"			/product="putative zinc finger protein"		
repeat_region	8658. .8684		misc_feature	/db_xref="PID:g3236261"		
	/rpt_family="AT-rich"			/translation="MFNNITTSFGSGGIVVVPPTATTPTTDFGTTTNSSTFIII		
	complement(9842. .9915)			GPPPPFPAPRSIDLTPLKLIFVIAFVAPALVAFNCGPSSSRNNSSTSS		
	/rpt_family=" (GA)n"			SDDTPHATVDPTTETVTSETSGGKFHKDTHSKETIGNECSVCLAVFTDSDLRLOSE		
misc_feature	10503. .10670		gene	CKHAFVLCIETNLKDHENPCICRDTVSQKQTEAPNVNVNNGSGNRVSAT		
	/note="exon predicted by xgrail, quality marginal_shadowexon"			SKDDWRGLSDASTVSVDKSDVDSQAQHTVIEDLILIGLRLLKAIHHPNRYL		
	10725. .10746			SFCSSKSDSPSLAPNRLNATKSTIPRFAVLVTGKDGKVRKRLKAIHHPNRYL		
	/rpt_family="AT-rich"			LHLDAESDERMELAKYVRGKKFENVMVGLADLVTEKGTPLASTLHGVAITLLK		
mRNA	complement(join(11416. .11526,11608. .12116,12551. .12628,12737. .13604))		mRNA	KAKDWMETINLSADYPLMPODDILHIESYLPYLNFTIETSNIGWKENQARPIIID		
	/gene="F13M22.5"			PGFYHLKSGYFWAKERSLPASFLMGSTVALTRPFLFPCINGWNLPTLLMY		
	complement(11416. .13604)			TNFLSSSGYFQTVVCNKDQNTVNDHLYTKWDLQORTLVNTVENFRDMVOSGA		
	/gene="F13M22.5"			PFARFERDDVLVDKIDIELLQOTDGLTELKTPDVVKPTVSWKRLEKLMVRLLDHEN		
gene	complement(join(11610. .12116,12551. .12628,12737. .13522))		CDS	/codon_start=1		
	/gene="F13M22.5"			/product="putative ARE1 GTPase activating protein"		
	/codon_start=1			/db_xref="PID:g3236238"		
	/translation="MAAARLRLTQSPENKVCVDCSQKNPQWASISYIGFMCLECSG			KIRGLGVHLSFVRSVTMDSWEIQIKMDAGNERLNLFQAIGISKETDIITSKNSN		
CDS	AASVYRDRJQALAEQRWDPTIVKESVGGGLMKNKPPLSQGGSDRGNGWDMND		mRNA	/db_xref="PID:g3236238"		
	DSFRSDTMRNOSAGDSRSGGAPAKSKSSEDIYSRQLEASANKKESFFAKRMAE			complement(join(11416. .11526,11608. .12116,12551. .12628,12737. .13604))		
	NESKBPGLPSOGKVYFGSPGAPRNSQSGGDVFSVMESEGFGRSLISAASAAAN			/gene="F13M22.5"		
	VYQTQMEFTSKVKEGLDQTVSETVNVVASKTTEIGORTGIMKGVMAIASQKVEEF			complement(11416. .13604)		
repeat_region	THEASTWNNQNKTENGYIQNSGINKTANSFCGSSQSSSGSHNSYRNSWDWG		gene	/gene="F13M22.5"		
	EENNSKKEAAPKYSTNSDDDDGGWAGWDNDKARDDDFYIOPASDKKSVGHNGKSDTA			complement(join(11610. .12116,12551. .12628,12737. .13522))		
				/gene="F13M22.5"		
				/codon_start=1		
repeat_region	/product="putative ARE1 GTPase activating protein"		CDS	/db_xref="PID:g3236238"		
	/db_xref="PID:g3236238"			/translation="MAAARLRLTQSPENKVCVDCSQKNPQWASISYIGFMCLECSG		
				KIRGLGVHLSFVRSVTMDSWEIQIKMDAGNERLNLFQAIGISKETDIITSKNSN		
				AASVYRDRJQALAEQRWDPTIVKESVGGGLMKNKPPLSQGGSDRGNGWDMND		
mRNA	DSFRSDTMRNOSAGDSRSGGAPAKSKSSEDIYSRQLEASANKKESFFAKRMAE		mRNA	/db_xref="PID:g3236238"		
	NESKBPGLPSOGKVYFGSPGAPRNSQSGGDVFSVMESEGFGRSLISAASAAAN			complement(join(11416. .11526,11608. .12116,12551. .12628,12737. .13604))		
	VYQTQMEFTSKVKEGLDQTVSETVNVVASKTTEIGORTGIMKGVMAIASQKVEEF			/gene="F13M22.5"		
	THEASTWNNQNKTENGYIQNSGINKTANSFCGSSQSSSGSHNSYRNSWDWG			complement(11416. .13604)		
repeat_region	EENNSKKEAAPKYSTNSDDDDGGWAGWDNDKARDDDFYIOPASDKKSVGHNGKSDTA		gene	/gene="F13M22.5"		
				complement(join(11610. .12116,12551. .12628,12737. .13522))		
				/gene="F13M22.5"		
				/codon_start=1		
repeat_region	/product="putative ARE1 GTPase activating protein"		CDS	/db_xref="PID:g3236238"		
	/db_xref="PID:g3236238"			/translation="MAAARLRLTQSPENKVCVDCSQKNPQWASISYIGFMCLECSG		
				KIRGLGVHLSFVRSVTMDSWEIQIKMDAGNERLNLFQAIGISKETDIITSKNSN		
				AASVYRDRJQALAEQRWDPTIVKESVGGGLMKNKPPLSQGGSDRGNGWDMND		
mRNA	DSFRSDTMRNOSAGDSRSGGAPAKSKSSEDIYSRQLEASANKKESFFAKRMAE		mRNA	/db_xref="PID:g3236238"		
	NESKBPGLPSOGKVYFGSPGAPRNSQSGGDVFSVMESEGFGRSLISAASAAAN			complement(join(11416. .11526,11608. .12116,12551. .12628,12737. .13604))		
	VYQTQMEFTSKVKEGLDQTVSETVNVVASKTTEIGORTGIMKGVMAIASQKVEEF			/gene="F13M22.5"		
	THEASTWNNQNKTENGYIQNSGINKTANSFCGSSQSSSGSHNSYRNSWDWG			complement(11416. .13604)		
repeat_region	EENNSKKEAAPKYSTNSDDDDGGWAGWDNDKARDDDFYIOPASDKKSVGHNGKSDTA		gene	/gene="F13M22.5"		
				complement(join(11610. .12116,12551. .12628,12737. .13522))		
				/gene="F13M22.5"		
				/codon_start=1		
repeat_region	/product="putative ARE1 GTPase activating protein"		CDS	/db_xref="PID:g3236238"		
	/db_xref="PID:g3236238"			/translation="MAAARLRLTQSPENKVCVDCSQKNPQWASISYIGFMCLECSG		
				KIRGLGVHLSFVRSVTMDSWEIQIKMDAGNERLNLFQAIGISKETDIITSKNSN		
				AASVYRDRJQALAEQRWDPTIVKESVGGGLMKNKPPLSQGGSDRGNGWDMND		
mRNA	DSFRSDTMRNOSAGDSRSGGAPAKSKSSEDIYSRQLEASANKKESFFAKRMAE		mRNA	/db_xref="PID:g3236238"		
	NESKBPGLPSOGKVYFGSPGAPRNSQSGGDVFSVMESEGFGRSLISAASAAAN			complement(join(11416. .11526,11608. .12116,12551. .12628,12737. .13604))		
	VYQTQMEFTSKVKEGLDQTVSETVNVVASKTTEIGORTGIMKGVMAIASQKVEEF			/gene="F13M22.5"		
	THEASTWNNQNKTENGYIQNSGINKTANSFCGSSQSSSGSHNSYRNSWDWG			complement(11416. .13604)		
repeat_region	EENNSKKEAAPKYSTNSDDDDGGWAGWDNDKARDDDFYIOPASDKKSVGHNGKSDTA		gene	/gene="F13M22.5"		
				complement(join(11610. .12116,12551. .12628,12737. .13522))		
				/gene="F13M22.5"		
				/codon_start=1		
repeat_region	/product="putative ARE1 GTPase activating protein"		CDS	/db_xref="PID:g3236238"		
	/db_xref="PID:g3236238"			/translation="MAAARLRLTQSPENKVCVDCSQKNPQWASISYIGFMCLECSG		
				KIRGLGVHLSFVRSVTMDSWEIQIKMDAGNERLNLFQAIGISKETDIITSKNSN		
				AASVYRDRJQALAEQRWDPTIVKESVGGGLMKNKPPLSQGGSDRGNGWDMND		
mRNA	DSFRSDTMRNOSAGDSRSGGAPAKSKSSEDIYSRQLEASANKKESFFAKRMAE		mRNA	/db_xref="PID:g3236238"		
	NESKBPGLPSOGKVYFGSPGAPRNSQSGGDVFSVMESEGFGRSLISAASAAAN			complement(join(11416. .11526,11608. .12116,12551. .12628,12737. .13604))		
	VYQTQMEFTSKVKEGLDQTVSETVNVVASKTTEIGORTGIMKGVMAIASQKVEEF			/gene="F13M22.5"		
	THEASTWNNQNKTENGYIQNSGINKTANSFCGSSQSSSGSHNSYRNSWDWG			complement(11416. .13604)		
repeat_region	EENNSKKEAAPKYSTNSDDDDGGWAGWDNDKARDDDFYIOPASDKKSVGHNGKSDTA		gene	/gene="F13M22.5"		
				complement(join(11610. .12116,12551. .12628,12737. .13522))		

```

repeat_region complement(20872..20971)
repeat_region /rpt_family="(TGG)n"
repeat_region complement(21518..21543)
misc_feature /rpt_family="AT-rich"
repeat_region complement(24200..24291)
repeat_region /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region complement(25078..25159)
repeat_region /rpt_family="(TA)n"
misc_feature join(<25358..25408,25622..>26563)
misc_feature /gene="F13M22.9"
gene <25358..>26563
CDS join(25358..25408,25622..26563)
CDS /gene="F13M22.9"
CDS /codon_start=1
CDS /product="putative zinc finger protein"
CDS /db_xref="PID:g3236241"
CDS /translation="MVFSIQAYLDSSNQAPPSSNYNHDGTGASANGHVLRLPOLP
OOQOOPHPNGSGGGGGGSIAGSNVDRARQANVALPEALKCPCESTNTKFC
YNNVSLTOPRHFCTCRYWRGGALRNPVGGGCRNRRTKSNNNNNSTATSN
TFSGNSATISTLSSHGGNQSILSQILSPARLMNPTYNHLGDLTSTNTKDNMS
LNYGLSODLSIRHAGSGGLMSCDWRASVHQQSGMGGNLSSNPNPANG
FYFSPRITSASISALASQFSSVKVEDNPYKVVNVNGNCSSWNDLSAFGSSR"
repeat_region 25691..25728
repeat_region /rpt_family="(CAG)n"
repeat_region 26002..26054
repeat_region /rpt_family="(CAA)n"
misc_feature complement(27038..27122)
repeat_region /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region 28179..28208
repeat_region /rpt_family="AT-rich"
misc_feature complement(28259..28328)
repeat_region /note="exon predicted by xgrail, quality
marginal_shadowexon"
repeat_region complement(28897..28980)
repeat_region /rpt_family="(TA)n"
repeat_region 29085..29133
repeat_region /rpt_family="AT-rich"
misc_feature complement(29945..30023)
misc_feature /note="exon predicted by xgrail, quality good"
misc_feature complement(join(<30476..30515,30659..30858,31168..>31269))
misc_feature /gene="F13M22.10"
misc_feature complement(<30476..>31269)
misc_feature /gene="F13M22.10"
CDS complement(join(30476..30515,30659..30858,31168..31269))
CDS /gene="F13M22.10"
CDS /codon_start=1
CDS /product="putative ribosomal protein L36"
CDS /db_xref="PID:g3236242"
CDS /translation="MTTPAVKTGLFVGLNKGHVVRRELAPRPNRSKGTSTRTIFIR
KLIREVAGMAPYEKRTITELLKVGDKRALKVAKRKLGTHTKRAKRREEMSSVLKMR
LGCAAAAEKKM"
misc_feature complement(31377..31401)
misc_feature /note="exon predicted by xgrail, quality marginal"
misc_feature complement(32006..32036)
misc_feature /note="exon predicted by xgrail, quality marginal"
repeat_region 32044..32161
repeat_region /rpt_family="(GAAA)n"
repeat_region complement(32282..32354)
repeat_region /rpt_family="(GAAA)n"
repeat_region 32653..32718
repeat_region /rpt_family="POLY_A"
repeat_region complement(join(<32935..33310,33630..>33739))
repeat_region /gene="F13M22.11"
repeat_region complement(<32935..>33739)
repeat_region /note="predicted by genscan"
repeat_region complement(join(32935..33310,33630..33739))
repeat_region /gene="F13M22.11"
repeat_region /note="hypothetical protein"
repeat_region /codon_start=1
repeat_region /db_xref="PID:g3236243"
repeat_region /translation="MEIGNKADVNLVNLREAERHRRYFDYFIDEILMWSFLGMS
SLKKEEVTLGKVKETRECESDDKLVLDICLKRRKRLDEEEVEEVEENDGFKTTRP
ENRIFIVRECPAPMKRSSEMFGRMTICRRRLSLFPEDDVNSFIDLOWRITTTTIK
K"
repeat_region complement(33140..33196)
repeat_region /rpt_family="(GAA)n"
repeat_region 33357..33398
repeat_region /rpt_family="(GAA)n"
repeat_region 33469..33489
repeat_region /rpt_family="AT-rich"
misc_feature complement(34248..34296)
misc_feature /note="exon predicted by xgrail, quality
marginal_shadowexon"
misc_feature 35206..35258
misc_feature /note="exon predicted by xgrail, quality good"
repeat_region complement(35273..35341)
repeat_region /rpt_family="(GAA)n"
repeat_region join(35801..35886,36023..36416,36517..37130,37242..37322)
repeat_region /gene="F13M22.12"
repeat_region 35801..37322
repeat_region /note="identical to GB:U29480"
repeat_region /gene="F13M22.12"
repeat_region join(35827..35886,36023..36416,36517..37130,37242..37307)
repeat_region /gene="F13M22.12"
repeat_region /codon_start=1
repeat_region /product="actin 3 protein"
repeat_region /db_xref="PID:g3236244"
repeat_region /translation="MADGEDIQPLVCDNGTMVKAGFAGDAPRAVFPISVGRPRHTG
VNVMGQDAYVGDEAQSGRGITLKYPIEHGIVNWDMEKIWHHTFYNELRVAPEE
HPILLTEAPLNPKANREKMTJMFETFNAPYVAIOAVLSYASGRTGIVLDSGDG
VSHYTVIYEGYALPHAILRLDLAGDLTALMKILTERGYSTTTAEREIVRDIKEKL
CYIALDYQELETAKTSSVKNYELPDGQVITIGSERFCEVLYQPSMIGMENAGI
HETTYSIMKCDVIDKLDYGNIVLSGGTTFPGIADRMSKEITLAPSSMKIKVVAP
PERKYSVMIGGSILASLTFQOMWIAKAEYDESGPSIVHRKCF"
repeat_region 37700..37817
repeat_region /rpt_family="(CAAAA)n"
repeat_region complement(<37864..>38967)
repeat_region /gene="F13M22.13"
repeat_region complement(<37864..>38967)
repeat_region /gene="F13M22.13"
repeat_region complement(37864..38967)
repeat_region /gene="F13M22.13"
repeat_region /codon_start=1
repeat_region /product="myb-related protein"
repeat_region /db_xref="PID:g3236245"
repeat_region /translation="MKQRWSGEDALLRAYVRQFQPREHVLVSERNKPLNRDAKS
CLERNKYLKPGIKKGLSTEEORLVIRLOEKHGNKKIAAEVPGRTAKRLGKWWEV
FKKQOREEKESKRVPEIDSKYDRILESFAEKLVKERSNVVPAATAATVVMANS
GGFLHSEQQVQPPNPVIPPWLATSNNGNVNVPSPVLTLSPSVAAAAPQPIPLW
QQQPERAENGPGGLVLSGMPSCSGSESFLSELCECCRELESGHRAWADHKEEA
WRLRLLEQLSEKTRQREKMEETEARKNALREQKNAMKEIEGEYREQLVGLRDA
EAKDQKLADQMTSRHRLTKFLEQQMGCLDRP"
repeat_region complement(39531..39561)
repeat_region /rpt_family="AT-rich"
repeat_region complement(40737..40818)
repeat_region /rpt_family="(TAA)n"
repeat_region complement(41089..41112)
repeat_region /rpt_family="POLY_A"
repeat_region complement(41190..41213)
repeat_region /rpt_family="POLY_A"
repeat_region complement(41283..41310)
repeat_region /note="exon predicted by xgrail, quality
marginal_shadowexon"
repeat_region complement(41558..41602)
repeat_region /rpt_family="AT-rich"
repeat_region complement(41966..42068)
repeat_region /rpt_family="(GAAA)n"
repeat_region complement(42281..42325)
repeat_region /rpt_family="AT-rich"
repeat_region 42889..42968
repeat_region /note="exon predicted by xgrail, quality excellent"
repeat_region 43132..43154
repeat_region /note="exon predicted by xgrail, quality good"
repeat_region 43371..43479

```

mRNA	/rpt_family="TAAAAA)n"	mRNA	complement(join(<53450..54129,54196..54289,54371..54453,54545..54596,54711..54868,54958..56602))
	/gene="F13M22.14"		/gene="F13M22.17"
	complement(43750..45963)		complement(<53450..>56602)
	/gene="F13M22.14"		/note="predicted by grail"
	complement(join(44143..44452,45240..45558,45719..45878))		/gene="F13M22.17"
CDS	/gene="F13M22.14"	CDS	complement(join(53450..54129,54196..54289,54371..54453,54545..54596,54711..54868,54958..56602))
	/codon_start=1		/gene="F13M22.17"
	/product="putative expansin protein"		/note="hypothetical protein"
	/db_xref="PID:g3236246"		/codon_start=1
	GGDSAGTMCACGCGNLYSGYGVNTAALSTALFNNGFCGACAEIKCDDPWCVP		/db_xref="PID:g3236249"
repeat_region	GNPSLVTATGCPNPNFAQSDDGWCNPPREHFDLAMPFLKLGIRAGLIVPSYRR	repeat_region	/translation="MEDGNKLDRRKMTMMWEGDGFEDDDDRFFETHDRLSALTFF
	VPCRKIGIRFTVNGFRYFNVLVTNAGADINGVSVKSGKTDVWRMSNMGWOS		DMANASSDEDEDPCRLSFSAVSSLTASTARKFTAMPSPDYDIWMAAPGISERR
	NAVIGQSLSRVTASDRSRSTSNVPATWQFGQTSGKNFRV"		RLLHGLASNDKMDVSAVIRRVVSNAPVVSNGEKKMKMKMNHVEQDQHDHPVPM
	44795..44903		LARSSEDIERFLIEKRKEELIGKISKORLTATYSTICTRICOYOTPTROTFA
	/rpt_family="AT_rich"		VCGKALRGGEALTSVMSNARVGAFFLIKNDTGKEFVNEYDEDMNRSLDLOT
repeat_region	46287..46312	repeat_region	CCGKLEEFKCVGSPVVKELMRNVRNINTEPLMDLRKFNYSLSKSVRLKRRGA
	/rpt_family="AT_rich"		ALLKNIKVAHSNLSRVADKVDSDTSPKKGDKHKHKGANEMVKVRPTGKYKELS
	complement(47301..47444)		ALHMCQIOAHEGAVMTIKFSODAHYLAGGADRVIHVWEVQCELSMNEGSLTPIH
	/note="exon predicted by xgrail, quality excellent_shadowexon"		PSLCDSGNEITVVEKKKGSGRRHHIPDYVHVPTVFSFSDKPVCSLKGHLDA
	excellent_shadowexon"		ILDLSKSKOLLSSMDKTVRLWDIETKTCLKLEAHNDVYTCIOFSPDENYFLSGS
mRNA	/rpt_family="AT_rich"	mRNA	LDKIRIWSIQDRHVWESDLHEMTAACYTPDQCGALIGSHKGICRAYTEDECKLSQ
	<48689..>50845		TNOIDVQSNKSKAKRKITSFQSPVNSPELVTSADSRILDDSEVIRHKFGFRNT
	/gene="F13M22.15"		CSQLSASYSQDGRIICASEDSQVYLWKNDFHRTSTLTQSHHEFHCKDVSAAVPMH
	<48689..>50845		GHYRGPPVPVQIHSKRRKRISTSQSPSTISSPTKEETSATGPTTSNRNKKSGLPFM
	48689..50845		PKKATKSQIQPEEAGPELGSEFSRSMNSEQHSRSEGESPSINTSRLSSWSWF
CDS	/gene="F13M22.15"	CDS	DSGGHGPTQIPTAMGWVIVTATVHGEIRSYQNEGLPRRIGRGTTLF"
	/codon_start=1		complement(56145..56194)
	/product="SCARECROW-like protein"		/rpt_family="(GAA)n"
	/db_xref="PID:g3236247"		56886..56744
	SSDHINTVVAAPENSVETREEEDPADDFDSDAVLGYSOMLNEEDDDKVCYMLQ		/note="exon predicted by xgrail, quality
repeat_region	ESLDEAERSLYEAIGKKYPPSPERNLAFERNSENLDVRVPGNYTGGCIGFNGG	repeat_region	marginal_shadowexon"
	IEANRFPENELINFRNCEIVSCSKSRSDRDEICEEERSKLPFAVGEDILRSD		complement(56743..56808)
	VVDKILVHPGSGESMFLNALDKKVEKKASDAQGGKRRRGRGRGRGGGG		/note="exon predicted by xgrail, quality marginal"
	QNGKVEVDLSRLIHCAQVAADRRCAQOLLKOIRLHSTPFGDGNORLAHCFANG		complement(57092..57128)
	EARLAVGSGIYKGVKPSRAAAVLKAHOLFACPPKLSYFITNKTIRLDVNSQ		/note="exon predicted by xgrail, quality marginal"
mRNA	AKLGVPPEYKAIAKDWAIOLEDIDRDEITVYVNCILYRAENLDESVKVQIRLQAY	mRNA	57429..57497
	NLNLGKPNDFVFGVINGAYNAPFTVTRFREALFHSFIDMLETIVPDEDERMFL		/note="exon predicted by xgrail, quality
	EMEVFGREALNVIACGEWRVERPETYKQHVAMRSGLVQVPDPDSIMKTSLKHVT		marginal_shadowexon"
	FYHKDFVTDQNRWLLOGKGRVTWALSVMKPEKA"		complement(57759..57834)
	49591..49709		/note="exon predicted by xgrail, quality excellent"
repeat_region	/rpt_family="(GAA)n"	repeat_region	complement(join(<57887..58388,59719..59846,60078..60132,60218..60291,60384..60527,61543..>61581))
	49635..49689		/gene="F13M22.18"
	/rpt_family="(GGA)n"		complement(<57887..>61581)
	51112..51155		/note="predicted by genefinder"
	complement(51301..51387)		/gene="F13M22.18"
misc_feature	/note="exon predicted by xgrail, quality good"	misc_feature	complement(join(57887..58388,59719..59846,60078..60132,60218..60291,60384..60527,61543..61581))
	/note="exon predicted by xgrail, quality		/gene="F13M22.18"
	marginal_shadowexon"		/note="hypothetical protein"
	complement(join(51428..51666,51782..51905,51988..52049,52131..52264,52306..52342,52427..52698,52779..>53043))		/codon_start=1
	/gene="F13M22.16"		/db_xref="PID:g3236250"
CDS	complement(51428..>53043)	CDS	/translation="MPRVVSVSTPAQAFSGTQNVNQOKEAMRVNVQIQGIDLEHG
	/gene="F13M22.16"		YLCGTMEALNVPNADPTVITFWEGEIVDKNITYFTGKWEATREDDMRHMSKPSFP
	complement(join(51547..51666,51782..51905,51988..52049,52131..52264,52306..52342,52427..52698,52779..53043))		LOWKQOYFVNVGTDGCLTAGFYVYCFSCDGSISGIFYDPNSFHHMIISSKNVL
	/gene="F13M22.16"		KMEVEVSKRKFTQSDDELSSLPLSKHTCFANVACSENTNGNSRIDTYSKSSVYN
	complement(51428..53043)		STTSMCNNDIENKSESSCGGDKMISFESHLDIYIGTONLEDFSEKVIENILYLDE
mRNA	/gene="F13M22.16"	mRNA	QEEEDDAKGCNNAARAFVLSGRWTVNOCKENRILV"
	complement(join(51547..51666,51782..51905,51988..52049,52131..52264,52306..52342,52427..52698,52779..53043))		58770..58795
	/gene="F13M22.16"		/rpt_family="(GA)n"
	/note="unknown protein"		59165..59234
	/codon_start=1		/rpt_family="AT_rich"
repeat_region	/db_xref="PID:g3236248"	repeat_region	59249..59312
	/translation="MAMMTTTTTTFFHPLLPANTYKSGAVASSVSVPRSSSQFRSL		/rpt_family="AT_rich"
	VSDSTICGSPSKFTGKNRRSVTVSAAATPELTVLTGAGRTGQIVYRKLKERSQ		complement(59446..59478)
	FVARGLVTKESKKEINGEDEVFIGDIRDTASIAPAVEGIDALVITSAVPMQKPGFD		/rpt_family="AT_rich"
	PSKGRPEFFDDGAYPEQVDWIGKQIDAGVDDGLSFVVTAKAGVKQIVLVGSM		complement(61713..61851)
CDS	GNTINISIGNANILLVWKRAEQVLADSGIYTIIRAGGLQDKGGRELLVGRD	CDS	/note="exon predicted by xgrail, quality good_shadowexon"
	DELLETETRTIARADVAEVCVQALQLEAKFKALDLASKPEGTGTPTKDFKALFTQVT		join(<62177..62406,62814..63035,63131..63242,63337..63399,63504..63642,63940..63995,64072..64145,64247..64394,64470..64652,64757..64844,64948..65058,65257..65354,
	TKF"		65354,

gene	65591..65708,65788..65882,66002..66108,66213..>66306) /gene="F13M22.19" <62177..>66306 /gene="F13M22.19"
CDS	Join(62177..62406,62814..63035,63131..63242,63337..63399, 63504..63642,63940..63995,64072..64145,64247..64394, 64470..64652,64757..64844,64948..65058,65257..65354, 65591..65708,65788..65882,66002..66108,66213..66306) /gene="F13M22.19" /codon_start=1 /product="putative phosphoribosylaminoimidazole carboxylase"
repeat_region	/db_xref="PID:g3236251" /translation="MLLKQSSAAVLVVGNTTPVLHTSRSTYRVCPFPVTQTQSFQSL TMANLQKLPITSSSKLNTASAPGSSHDASISENRKNHVGSEIKRVGLGGQLG RMLQOASQLAKVMDPSKNSASALSYGHMVDSDFSATVEEFKAKVGLGGQLG HVDYTLKLEKQCPQPKASTIRIQDKYMKVHFSGHGPILPEFMEISIDEGAK AGELFYPLMIKSKRLAYDGRGNANQDELSSAVTALGGFSRGLYTEKWAIPVKEL AVIVARGDSMVCYPVYETIHRNCHIVKAPADVPWKINKLATDVAQKAVGSLEGA GVFAVELFTSDSQILLNEVAPRHNSGHQIECCYTSQFQHLRAVYVPLGDPFSHR TPASIMNILEDDVIDGEAGFKLAHRLIARALCPGASVHWYDKPERMKRKGHIT LVGQSMGLEPRLQCIIESEQSHQVETPRVAIIMGSDTDLPMKDAKILDLFGVTHE KIVSAHRTPEMPTYATSAHSRGVQVITAGAGAAHLPGMVASLTPLPVTGVPRAT RLDGVDLSLSTVQMPRGVPVATVAINNATNAALLAVRLMIGISDSDLVSRMQYQEDMR DENLNGKLEKTEGWSYLNQ"
misc_feature	complement(join(<66647..66829,67056..67229,67307..67405, 67776..67976,68080..68187,68273..68651,68917..69175, 69620..70064,70183..>70242)) /gene="F13M22.20" complement(<66647..>70242) /gene="F13M22.20"
repeat_region	/gene="F13M22.20"
repeat_region	complement(join(66647..66829,67056..67229,67307..67405, 67776..67976,68080..68187,68273..68651,68917..69175, 69620..70064,70183..70242)) /gene="F13M22.20"
misc_feature	/codon_start=1 /product="CER1-like protein" /db_xref="PID:g3236252" /translation="MASRPGFLDMPWTPLGSFKYLLAPLVFDSIYSYATIRDEKHL LIVATVYVHISQWISLSRYQAKGTRILKNSIEFDQVDRERTWDDQIFNTLIV YLTVYVSGTIPFWRDGVILVLAHAGVPEFYIYVFWRALHHHLYSHYSHSHHS SIVTEPITLCATNSKPVLIIVAVVHPFAEHIGYTLIGLPLITTFMCTGVSVSIALY LTYIDFNMGCHGFELPKFELSLLPPLKCYTPSFHSLHHTQFRTNYSLEFPMYD YIYGTDECSSLSLETSLEKEERDPAIHLLHLISLDSIYHLRGLFASLSHSPUSSRC YFLMKFPALLISFLRSFSQFTFVWRNRFRDLTHSHLLPKFSSHYMSHQKECIN KMIIPAALEADKGVKMSGLNQLNGYGEYVRRHPKLRIRIYDVGSLAAEVY LHSIPVGTKEVLRQKITKVARAIVSLCQNAIKVMVLRKEEHSMLAEFLDDCKENL VLTNNYPMIWLGDGLSTKEQMAKDGTLFLPFSORPPKTLRKDCFYHTTPAMTIPH SAQNIQSCENLGRVMSARVGGIVHALEGWKEHCEGLDDNSIINPRVWEALRNG FQPLLLESLEN"
repeat_region	complement(67680..67715) /rpt_family="AT_rich"
misc_feature	complement(70565..70634) /note="exon predicted by xgrail, quality marginal"
misc_feature	complement(70808..73053) /gene="F13M22.21"
repeat_region	complement(70808..73053) /note="highly similar to GB:X95909(polymorphism exists between nucleotides 73005 and 72085 which is a GA repeat. We found 6 copies in our sequence whereas only 5 copies exist in GB:X95909)" /gene="F13M22.21"
misc_feature	complement(71000..73027) /gene="F13M22.21"
repeat_region	/codon_start=1 /product="receptor-like protein kinase"
misc_feature	/db_xref="PID:g3236253" /translation="MFLKLLTIFFTFFNLIQSSQSLNFAYNNGFNPPDTLSIQGI TVPNGLLKTNTVQTKTHAFYTKPIRFKDSNPTGVSSFTSFVAIHQIAILSG HGIAFVAPNAPSLPQGNPSQVIGLFLNANNGETNHPVAVELDITILSEFNDNHNH GIDINSLKSVQSPAGYWDKQGNLTLISRKPMQVWDYDGRNKTIDVTMAFNPED KPTRLPVTAVRDLSSLQDMYGVFSSATGSLSEHYILGHSFGLNEKAPLALSRLP KLPRFPARISEFYKIGMPLISLILFISFIFLVCIYVRRRRKFAEELEWEKEFGKNR
repeat_region	complement(73447..73473) /rpt_family="AT_rich"
misc_feature	join(<74148..74206,74511..74566,75952..>77285) /gene="F13M22.22"
repeat_region	<74148..>77285 /gene="F13M22.22"
repeat_region	join(74148..74206,74511..74566,75952..77285) /gene="F13M22.22"
repeat_region	/note="unknown protein"
repeat_region	/codon_start=1 /db_xref="PID:g3236254"
repeat_region	/translation="MVQIVPVANAHMVKTSRKGLLALVVVGVVECCPTRCDLRDPN LVDSNCLSRSTHQSLEDVSIINHFRLNFRNRRASIFLCRLIGVICSQSKMYGG KSYTLKQSVLSLILILVTLVSAKNQPRI SLIEVKTCLNLAGWVEDKRPPLYS GFYTLQVSNLFINFCRVGRPDSEFEGYRWQPCGNIPEFNRVNFRRMONKTIAGIGD SLGRQFQSLMCMATGKESPEVONVSGEGLVTPKGPARGGNAYRFPNTNTVLSY WSALQDLVPMNNTDPHILIAMHLDRPAPFIRNLYLHRFVLVLTGHHWSRDXLEKNH WYMHVNTRVGGYFKNVENAKIETIISLWKVMDAQLPLHPLKAFITTIISPRHEKN YTIPLSRGSKITGEGSLDTIVESAVNGTRVKILDTALSRLRDEAHGACGLKPKKA SNVTSAPFNDCLHWCPLGPDPTWNELLIAQL"
repeat_region	complement(75409..75618) /rpt_family="AT_rich"
repeat_region	complement(77388..77424) /rpt_family="AT_rich"
repeat_region	complement(77464..77524) /rpt_family="AT_rich"
repeat_region	77550..77635 /note="exon predicted by xgrail, quality excellent_shadowexon"
repeat_region	complement(join(<78252..78416,78604..79152,79555..79945, 80111..>80604)) /gene="F13M22.23"
repeat_region	complement(<78252..>80604) /note="predicted by gensec"
repeat_region	complement(78252..78416,78604..79152,79555..79945, 80111..80604)) /gene="F13M22.23"
repeat_region	/note="hypothetical protein"
repeat_region	/codon_start=1 /db_xref="PID:g3236255"
repeat_region	/translation="MAIFDPFKLHKPPSIPATTVKPSHVLVLVAKFFITICIFISV AMTSFYIFFGSCDCHITGLRRLGPENVNLTNNSASSSTIRRPNOSSEATDISHIA FGIGSITQWRDRSRYSELMWRPNVTRGFIWLDEPPLNMTLSTSPYQVSAOTSRI SYCWYGSRAIMARIKETFGELTDVRWFIMGDDTVFVDNLITLVLNKYDHNOM YIIGNSSEVEDIVHAYMAYGGGTAISYPLAVELVLLDGDIDRYASLYSDOKI EACLSGVLTKELGPHQVDIRNGPIGLAAHPVALVLLHLDYVDPFPPTQID ALRLVSAIKTDPRIIOHSEFCHDQTRNWNYSWGYTIQIYTLVTAKELETPLTF KSWRTSSPEFSDTRISEDCPCERPLVFLDRVYEVGSGQTLTKRKHVEGSESTQC NSPDYSRANPEFIDVSTTTLTLPDKMAPRQCCEIVNSEDESSESVINVKIRHNPL ESVTPQSKHPNDEWTSVLSTETI"
repeat_region	complement(79324..79361) /rpt_family="AT_rich"
repeat_region	80698..80727 /rpt_family="AT_rich"
repeat_region	complement(81151..81260) /note="exon predicted by xgrail, quality excellent"
repeat_region	82037..82089 /note="exon predicted by xgrail, quality excellent"
repeat_region	/note="exon predicted by xgrail, quality excellent_shadowexon"
repeat_region	82508..82558 /note="exon predicted by xgrail, quality marginal_shadowexon"
repeat_region	complement(<83772..>84686) /gene="F13M22.24"







e-mail: rounsley@tigr.org  
BAC clone fl3M22 is from Arabidopsis chromosome II and is near the molecular marker ve018.  
The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge).  
<http://gnomic.stanford.edu/~chris/genscanw.html>, and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetpgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at/at.html>).

Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

## FEATURES

```

source
    Location/Qualifiers
    i. 93845
    /organism="Arabidopsis thaliana"
    /cultivar="Columbia"
    /db_xref="taxon:3702"
    /chromosome="II"
    /map="ve018"
    /clone="F13M22"
    complement(10..29)
    /note="exon predicted by xgrail, quality marginal"
    168..249
    /note="exon predicted by xgrail, quality good_shadowexon"
    <464..>619
    /gene="F13M22.1"
    <464..>619
    /note="predicted by genscan"
    /gene="F13M22.1"
    <464..>619
    /gene="F13M22.1"
    /note="hypothetical protein, 5' partial"
    /codon_start=1
    /db_xref="PID:g3236260"
    /translation="SRGVRVQRTHKDGDKRLDPATRRKLETNALIASGKGF
    EIGNGL"
    complement(Join(835..1212,1373..1513,1600..1693,
    1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
    2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
    4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
    /gene="F13M22.2"
    complement(835..>5654)
    /gene="F13M22.2"
    complement(Join(1099..1212,1373..1513,1600..1693,
    1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
    2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
    4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
    /gene="F13M22.2"
    /note="unknown protein"
    /codon_start=1
    /db_xref="PID:g3236235"
    /translation="MPKENGQLKRLDDDDTEGNGKDHFFSKKQAKASNDITISE
    INVPASVLESLFDVSSQPVKGLVCECSDFGSETVSDASVGSSTQESSD
    VLPSEVLVPHLSTGITKITFLSKPKKFDLPLIKDHTDAGVYKMPKKTIVS
    LSPSVNKLLETGILEGARVKYISTPPVROLLGIHSGVLCGCTTCNFSKVLISAYE
    FQHCACATRHNNHIFLENRAVYNIVOELKTPAPRVLEEVIRNVAGSALNEEGLRA
    WQGDGQVSNRYTACTGILRIIPLSVNALLSHYFVLSETDSCSCSYLPGFLDES
    QSLTPCSVNHVSEFSEKTDLEPKRIAKLTSHVSGTGCHKVKVSEGNRRKNDL
    HRLLEMPNGLPDGTLAYVVKTIQSPSQFEAHAGMAARQRYRHIFISSGLSLHDIAM

```

```

SLANGHVTTGSDSDMCICDGGDLLCAGCQPAHFACILFQSMPEGTWCSCND
GPISKATTTDPGNARPIVIRLSRVVKAFESDIGGVCFCRSNHSDFSGKFDRTVIL
CDOQKEVHVGCLRENGCDLKEIPOEKWFCSCNCRITHAVONSVCSPOTLPTPL
DMICKDREKGIPTDGTWIRILSGKSRYPHEHLPLLSRAAIFRECDFDPIVAKSGR
DLIPVMYGRNISGOEFGMCLVLIVNSLVYSAALLRFGVEAELFIVATSRIGQ
RGYFQGLYACVENVLLSNVENVLPLPAEAEASIWTKKFGTKMSDQOLQYQKEVQL
TIFGKTSMLKKVKPATTGLSESTLI"
6010..6135
/rpt_family=" (GGA)n"
6023..6202
/note="exon predicted by xgrail, quality
marginal_shadowexon"
6179..6236
/rpt_family="POLY_A"
Join(<6647..6702,7161..7304,7322..>7499)
/gene="F13M22.3"
<6647..>7499
/note="predicted by genscan"
/gene="F13M22.3"
Join(6647..6702,7161..7304,7322..7499)
/gene="F13M22.3"
/note="hypothetical protein"
/codon_start=1
/db_xref="PID:g3236236"
/translation="MTWKVREENKSFQFLFKPAMRIPHFVYGATWTLFLTLVAII
SLAPEFAVSALFSSSEVSFRRHIGSAGSSGLGVFTGESPEVEDGPRCAAGFCN
CCGFRRRRDYGLGRRSPLI"
complement(Join(<7761..7931,8019..8065,8154..8238,
8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
/gene="F13M22.4"
complement(<7761..>9886)
/gene="F13M22.4"
complement(Join(7761..7931,8019..8065,8154..8238,
8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
/product="putative ribitol dehydrogenase"
/codon_start=1
/db_xref="PID:g3236237"
/translation="MGIIGVMTGKKKGSGFGSASTAEDVTQADASHLTAITITGGTSG
IGLEAARVLMRGHVIIAARNPKAENSKEMILQMPNARYDIQIDVSSIKSVRSF
VDQFLANVPLNINAGVMFCFKLTEDGIESQFATNHIGHLTLNLLDKMSTA
RESGVQGRIVNLSIAHTYVSEGIKINDPAGYSERRAYGQSKLSNLLHNSALSR
DLOEGVNTITNSVHPLVNTNLFYSGFSKMKVRAMTFLWKNIPQGAATTCYVALII
PDLEGVTKYFGDCNIVAPSKFATNNSLADKLWDFSVFLIDSISR"
8658..8684
/rpt_family="AT-rich"
complement(9842..9915)
/rpt_family=" (GA)n"
10503..10670
/note="exon predicted by xgrail, quality
marginal_shadowexon"
10725..10746
/rpt_family="AT-rich"
complement(Join(11416..11526,11608..12116,12551..12628,
12737..13604))
/gene="F13M22.5"
complement(11416..13604)
/gene="F13M22.5"
complement(Join(11610..12116,12551..12628,12737..13522))
/gene="F13M22.5"
/codon_start=1
/product="putative ARF1 GTPase activating protein"
/db_xref="PID:g3236238"
/translation="MAAARLRLTQSPENKVCVCSOKNPQWASISYGFMCLECSG
KRGJGVHISFVRSVTMDSWSEIQIKKMDAGNERLNFLAQYISKETDIISKYNSN
AASVTRDIRIALAERQNRDPPIVRESVGGGKMMKPPPLSQSGGRSGNGGNDWND
DSFRDTRNRNQSAGDFSSGGGAPAKSDIEDYSRQLEASANKRESFFAKRAAE
NESKPEGLPPSQGGYVGFSSGPAPRNQSGGDFVSVMESEGFRLSLVSAAN
VYOTGTMEFTSKVKEGGLDVTSETVNVVASKTTEIGORTWIMKGVMAISQKVEEF
TKEASTNNQNKTEGNYONGSIGNTANSFGSGSSSSGHNSYNSNSWDDWG
ENNSKKEAAPKPVSTISNDDDDGGWAGDNDKADDDFFYQPADSKSKSVCHNGKSDTAW
TGGGFL"
11697..11757
repeat_region
misc_feature
repeat_region
mRNA
gene
CDS
repeat_region
misc_feature
repeat_region
mRNA
gene
CDS
repeat_region

```

```
/rpt_family="(CAT)n"
13716..13952
/notes="exon predicted by xgrail, quality
excellent_shadowexon"
complement(14172..14309)
/notes="exon predicted by xgrail, quality marginal"
join(14707..14911,15566..15765,15858..15964,16057..16189,
16311..16602,16755..17294)
/genes="F13M22.6"
14707..17294
/notes="identical to GB:U40269"
/genes="F13M22.6"
join(15634..15765,15858..15964,16057..16189,16311..16602,
16755..17182)
/genes="F13M22.6"
/codon_start=1
/product="origin recognition complex protein"
/db_xref="PID:g3236239"
/translation="MED1ENLEDEYGFERNYFLAKELGGASKSAHKLSDIHIVDEQ
ELRETASTIENKHSKEISELSMDKTMYSKWFLRCFGGLMTFGSKKALVEDFAS
ASLDYISVVINGILPSYLNKOVLLAELSELKCKRKSGLSGKQETFPERSMD
DILSLHPGQGDKDCFCVYVHNIDGPAIDPESQOTLARLSSCHIRLVAIDHVN
APLWDKKWIKQFNLMHVPTEAPYVNEGVTPLVLAOGSTAQTAKTAIVQSILT
PQNQVFKILAEYOLSHDEGDMPTDDLYSASREFVSSQVTLNLSHLTEFKDELVK
TKRNSDGECLNIPLTSAIRQLLDLQ"
complement(<17532..>19016)
/genes="F13M22.7"
complement(<17532..>19016)
/genes="F13M22.7"
complement(17532..19016)
/genes="F13M22.7"
/notes="unknown protein"
/codon_start=1
/db_xref="PID:g3236240"
/translation="MENHPSTLLSMOSSASHEELDMMNNRQSLLSGPPDNLPL
SAERPPPPNPLDAGCDLDVGLGQAVETENYMSVVKYGRKCAKRVDSWGAWFF
SFYKPKALNEKSAKIVRDSNGISGFQKDLKLDVLYOHDHNMWYKVRERENAL
GKMQRSTWNGSKQGRDLFFSVKEGFRVRMRQKRYHGLSNPQCQVHGLELPLN
LTKLDEERKRWELTGDLNFTIPPEASDPGSRNLPNTDFELERTPSLKNPTANH
SKLLNGSLNLOPSHNGEATDISPSHKKRKLDFNSGIIHEECCLTNVPOPPV
IEAQNELPTNEFTGAMKNVGPVTAAKTIYEDEGYLIIISLPFVDLSVKSVR
NTLTHGIKVSCLSTRVPFTKRHRDRTFKLTDSEHCPCGFEVREIPLSNRIPEDAN
IEAYDGGSGVLEILVPLKLRAGPEHEVRVCLRNUGNDLML"
complement(19045..19091)
/notes="exon predicted by xgrail, quality marginal"
19455..19545
/notes="exon predicted by xgrail, quality
excellent_shadowexon"
join(<20811..21507,21667..21805,22028..22434,22931..23006,
23118..23233,23331..>23788)
/genes="F13M22.8"
<20811..>23788
/genes="F13M22.8"
join(20811..21507,21667..21805,22028..22434,22931..23006,
23118..23233,23331..23788)
/genes="F13M22.8"
/codon_start=1
/product="putative zinc finger protein"
/db_xref="PID:g3236261"
/translation="MENNTTSFGSGGIVVVPATVPTDTPFGTTITSNSTFIIL
GPPPPFAPPSIDLTLKLIFFVIAFVAPVALYALFNGCPSRRNSSSRSTSS
SDDPFATVPDTPITETVTSSEGGKFHKDTHSKEIGNECVSLMFTDSDLEQLSE
CKHAFVLCIETLKHDPNCPICRDTVSQKQTEAPNVPVNVNGVNRSGNRVSAT
SRDDWRQGLPDASTVTSVRDQSDVQQAHTVYEDLILIGRLLIHLRLPIRL
SFCRSKPSDSSGLAPNRNLATKSTIPREAVLTGTGDKRVKRLKATHHPNNYL
LHLDLSEDERMELAKYVRSEKKFENVMVGLADLYTEKPTMLASTLHGVAILLK
KAKDWNFINUSADYFLMPQDDILHIFSYLPLRYLNFTEHTSNIGWKENQRARPIID
PGFVHLKRSQVFWAKERSLPASFKLPMGSTVALTRPFLFCIWNQDNLPTLLMY
TNFLSSEGQFYVVCNKNDYQNTVNHDLHTYKWDPLQOQRTLVNIVENFRDMVQSGA
PFAREFREDDLVLDKIDIELLQDQDTGLTKLTPDVVKPTVSWKRLKLMVRLDHENF
RAKOCK"
complement(20872..20971)
/rpt_family="(TGG)n"
complement(21518..21543)
/rpt_family="AT-rich"
complement(24200..24291)
/notes="exon predicted by xgrail, quality good_shadowexon"
complement(25078..25159)
/rpt_family="(TA)n"
join(<25358..25408,25622..>26563)
/genes="F13M22.9"
<25358..>26563
/genes="F13M22.9"
join(25358..25408,25622..26563)
/genes="F13M22.9"
/codon_start=1
/product="putative zinc finger protein"
/db_xref="PID:g3236241"
/translation="MVFSSIOAYLDDSSNQOAPPSNYNHDGTGASANGHVLRLPOLP
QQOQOQPHNGSGGGGGGSTRAGSMYDRAQANVALPEALKCPCESTNTKFC
YFNYSUTQPHRCKTCRRYWRGGALRNVPVGGCRRNRRTKSNNNNNSTATSN
TFSNGASTISTILSHYGNQESILSQILSPARLMPNTYHLDLTSNTKTDNNMS
LLNYGLSQDLRSIHMAGSGSLSCYDEWRSASYHQOSMGNGNLEDSPNPSANG
FYFSPRITSASISSALASQFSVKVEDNPYKVVNVNGNCSSWNLDLSAFGSSR"
25691..25728
/rpt_family="(CAG)n"
26002..26054
/rpt_family="(CAA)n"
complement(27038..27122)
/notes="exon predicted by xgrail, quality good_shadowexon"
28179..28208
/rpt_family="AT-rich"
complement(28259..28328)
/notes="exon predicted by xgrail, quality
marginal_shadowexon"
complement(28897..28980)
/rpt_family="(TA)n"
29086..29133
/rpt_family="AT-rich"
complement(29945..30023)
/notes="exon predicted by xgrail, quality good"
complement(join(<30476..30515,30659..30858,31168..>31269))
/genes="F13M22.10"
complement(<30476..>31269)
/genes="F13M22.10"
complement(join(30476..30515,30659..30858,31168..31269))
/genes="F13M22.10"
/codon_start=1
/product="putative ribosomal protein L36"
/db_xref="PID:g3236242"
/translation="MTTPAVKTGLFVLNGHVTVTRRELAPRPNRSKRGKTSKRTIFIR
KLIREVAGMAPYKRIITELDLKVKRALKVAKRKLGTGTHKRAKREEMSSVLKRWRS
LGGAAAAEKKM"
complement(31377..31401)
/notes="exon predicted by xgrail, quality marginal"
complement(32006..32036)
/notes="exon predicted by xgrail, quality marginal"
32044..32161
/rpt_family="(GAAA)n"
complement(32282..32354)
/rpt_family="(GAAA)n"
32653..32718
/rpt_family="POLY-A"
complement(join(<32935..33310,33630..>33739))
/genes="F13M22.11"
complement(<32935..>33739)
/notes="predicted by genscan"
/genes="F13M22.11"
complement(join(32935..33310,33630..33739))
/genes="F13M22.11"
/notes="hypothetical protein"
/codon_start=1
/db_xref="PID:g3236243"
/translation="MEIGNKADVNLVNLREARHRRYFDYIDEILWLSFLGMSSET
SLAKEEYTLGGVKETREECSDDKLDICLKRKLDEEEEEEVEEENDGFKTPRP
ENRIPVRECAPPAPMKRSSEMFGRMTMYCRRRLSFLPEDDVNSFITDLQWRTTTIK
```

repeat_region	K <sup>a</sup>	complement(33140. .33196)	gene	/gene="F13M22.14"
repeat_region		/rpt_family="(GAA)n"		complement(43750. .45963)
repeat_region		33357. .33398	CDS	/gene="F13M22.14"
repeat_region		/rpt_family="(GAA)n"		complement(join(44143. .44452,45240. .45558,45719. .45878))
misc_feature		33469. .33489		/gene="F13M22.14"
misc_feature		/rpt_family="(GAA)n"		/codon_start=1
repeat_region		complement(34248. .34296)		/product="putative expansin protein"
misc_feature		/note="exon predicted by xgrail, quality		/db_xref="PID:g3236246"
repeat_region		35206. .35258		/translation="MTATPRVGLWLTAVTASFLLTATNAKIPGVYSGPWNHATFV
repeat_region		/note="exon predicted by xgrail, quality good"		GGDSAGTMGACGYNLYSQGYGNTAAALSTALFNMGFCGACFEIKCTDDPRWCVP
repeat_region		complement(35273. .35341)		GNPSILYATNFPCPNFAQSDGWCNPPREHFDLAMPFLKIGLRAGIVPVSYRR
repeat_region		/rpt_family="(GAAA)n"		VPCRKIGGIRFTVNGFRYFNLYLVNTVAGADINGVSGKSTKDVMSRMWGNWQS
repeat_region		join(35801. .35886,36023. .36416,36517. .37130,37242. .37322)	repeat_region	NAVLIGQSLSRVTSADRRSSTSNVAPATWQFGQTFSGKNFRV"
repeat_region		/gene="F13M22.12"		44795. .44903
repeat_region		35801. .37322		/rpt_family="AT_rich"
repeat_region		/note="identical to GB:029480"	repeat_region	46287. .46312
repeat_region		/gene="F13M22.12"		/rpt_family="AT_rich"
repeat_region		join(35827. .35886,36023. .36416,36517. .37130,37242. .37307)	misc_feature	complement(47301. .47444)
repeat_region		/gene="F13M22.12"		/note="exon predicted by xgrail, quality
repeat_region		/codon_start=1		excellent_shadowexon"
repeat_region		/product="actin 3 protein"	repeat_region	48559. .48589
repeat_region		/db_xref="PID:g3236244"		/rpt_family="AT_rich"
repeat_region		translation="MADGEDIQPLVCDNGTGMKVRAGFADDPRAVFPISIVGRPRHTG	repeat_region	<48689. .>50845
repeat_region		VMVGMGQDAVGDQAQSKRGILTLKYPLEHGIVNWDDEKIKWHHTFYNELRVAPEE	repeat_region	/gene="F13M22.15"
repeat_region		HPILLTEAPLANKREKQIMETFNAPMYAIOALVLSYASGRTTGIVLSGDC	repeat_region	<48689. .>50845
repeat_region		VSHVPIVEGALPHAILRLDLAGRLDLMKILTERGYSFTTAEIRIVRDLKERL	repeat_region	/gene="F13M22.15"
repeat_region		HYTALDYELELTAKTSSVKNYELPDQVITIGSERFCPEVILYOPSMIGMENAGI	repeat_region	48689. .50845
repeat_region		CETIYNSIMKCDIVRDKLYGNIVLSGTTMFPGLIADRSKEITLAPSSMKIKVAP	repeat_region	/gene="F13M22.15"
repeat_region		PERKYSVWIGGISLASLSTFOOMIAKAEYDESGSIVHRKCF"	repeat_region	/codon_start=1
repeat_region		37700. .37817		/product="SCARECROW-like protein"
repeat_region		/rpt_family="(CAAAA)n"		/db_xref="PID:g3236247"
repeat_region		complement(<37864. .>38967)	repeat_region	/translation="MITEPSLTGISGMVNRNLSGLPDQSSHSFTPTVLYDGFNYNL
repeat_region		/gene="F13M22.13"		SSDHINTVVAAPENSIVIRIEEEDPADDFDSDAVALGYISQMLNEEDMDKVCMLQ
repeat_region		complement(<37864. .>38967)		ESLDLEAASLYEAIKGGKPPSPERNLAENSENLDRVYPNGYTGDCGCGFNGG
repeat_region		/gene="F13M22.13"		IKPLSSGFTLDRNPQSCSSILSVQSNGLITTYDGDIDESKNRNHQVWLFRR
repeat_region		complement(37864. .38967)		IEEANREPNELIVNFRRENCVSKARKNSRDEICVEEERSKLPVFGEDILRSD
repeat_region		/gene="F13M22.13"		VVDKILVHPGGESMKFENALRDVLKKGVEKKASDAQGKRRARGRGRGGGG
repeat_region		complement(37864. .38967)		ONGKKEVDLRSLLIHCAQAVAADRCAGOLLQIRLHSTPFGDGNORLAHCFANGL
repeat_region		/gene="F13M22.13"		EARUAGISQYIKGIVSPSAALVLAHQFLACCPFRKLSYITFNKTLINDLVGNSQ
repeat_region		/codon_start=1		RHVHDFGILYGFQMTLIRHFSVMYSPKVRITIEFPQGFQPRFAVEETGQRLAAY
repeat_region		/db_xref="PID:g3236245"		AKLFGVPFEYKAIKAKWDALQLEDLDRDEITVYVNCILYRAENLHDESVKVESCRDTV
repeat_region		translation="MKERQSRGEEDALLRAYVRQFGPREHVLVSERNKPLNRDAKVS		LNLIKGNPDLVFEVINGAYNAPFVTRFEALFHTSSIFDMLETIVPREDERMFL
repeat_region		CLEWKNYLKPQIKKSGLTSEEORLIVRLEKHGKWKKIAEYPGRTAKPLNGKWEV		FYHKDFVIDDNRMLQGWKQRTVYALSVWKPESKA"
repeat_region		FKEKQREKESNRKVEPIDEKSDRILESPFAEKLVKERSNVVFAAAAAAVVWANSN	repeat_region	49591. .49709
repeat_region		GGFLHSEQQVPPNPVIPPWLATSNNGNVNVARPSVLTLLSPSTVAAAAAQQPPPIPL		/rpt_family="(GAA)n"
repeat_region		QOQPERAENSGPGLVLSGMPSCSSSESVFLSELVECCRELEEGHRAWDHKKEA	repeat_region	49635. .49689
repeat_region		WLRRLLEQLSEKTCRQREKMEIEAKMKALREQKNAMEKIEGEYREQVLGLRRDA		/rpt_family="(GGA)n"
repeat_region		EAKDQKADQWTSRHIRLTKFLEQQMGCRDRP"	repeat_region	51112. .51155
repeat_region		complement(39531. .39561)		/note="exon predicted by xgrail, quality good"
repeat_region		/rpt_family="AT_rich"		complement(51301. .51387)
repeat_region		complement(40737. .40818)		/note="exon predicted by xgrail, quality
repeat_region		/rpt_family="(TAA)n"		marginal_shadowexon"
repeat_region		complement(41089. .41112)		complement(join(51428. .51666,51782. .51905,51988. .52049,
repeat_region		/rpt_family="POLY_A"		52131. .52264,52306. .52342,52427. .52698,52779. .>53043))
repeat_region		complement(41190. .41213)		/gene="F13M22.16"
repeat_region		/rpt_family="POLY_A"		complement(51428. .>53043)
repeat_region		complement(41283. .41310)		/gene="F13M22.16"
repeat_region		/note="exon predicted by xgrail, quality		complement(join(51547. .51666,51782. .51905,51988. .52049,
repeat_region		marginal_shadowexon"		52131. .52264,52306. .52342,52427. .52698,52779. .53043))
repeat_region		complement(41558. .41602)		/gene="F13M22.16"
repeat_region		/rpt_family="AT_rich"		/note="unknown protein"
repeat_region		complement(41966. .42068)		/codon_start=1
repeat_region		/rpt_family="(GAAAA)n"		/db_xref="PID:g3236248"
repeat_region		complement(42281. .42325)		/translation="MAWMTTITTTFFHPLLPANTYKSGAVASSFVSVPSSSLOFRSL
repeat_region		/rpt_family="AT_rich"		VSDSTICGPKSGTKNRRSVTVSAAATPEPLVLTGAGGTGOIVYKCLKRSEQ
repeat_region		42889. .42968		FVARGLVTRKESKEKINGEDEVFIGDIRDTASIAPEVIGDALVILISAVPQMPGED
repeat_region		/note="exon predicted by xgrail, quality excellent"		PSKGRPEFFDDQAPQEVQWICQKQIDAGVDDGLSVVTKARAGVKRQIVLWSM
repeat_region		43132. .43154		GGTNINPLSIGNANILVWKRKAQYLDASGIPITYIRAGLQDKDGGIRIELVSGD
repeat_region		/note="exon predicted by xgrail, quality good"		DELLETETRTIARADVAEVCVQALQLEAKFKALDLSKPEGTGTPTKDFKALFTQVT
repeat_region		43371. .43479		TKF"
repeat_region		/rpt_family="(TAAAA)n"		complement(join(<53450. .54129,54196. .54289,54371. .54453,
repeat_region		complement(join(43750. .44452,45240. .45558,45719. .45963))		54545. .54596,54711. .54868,54958. .>56602))

```
gene
/genes="F13M22.17"
complement(<53450..>56602)
/notes="Predicted by graal"
CDS
/genes="F13M22.17"
complement(join(53450..54129,54196..54289,54371..54453,
54545..54596,54711..54868,54958..56602))
/genes="F13M22.17"
/notes="hypothetical protein"
/codon_start=1
/db_xref="PID:g3236249"
/translation="MEDGNLDRKTKTMNWEGLGDFEDDDDRFFETHDRLSALTF
DMANASSDSEDEDFDCRLSFSNAVSLTTASRKERTPAMSPDYDIMAAPQGISERR
RLRHGMLGNMDMSAVSIRRVSNAPVVSNGEKKKKKKNNHNEQDHDHPVPM
LARSSESDIERELIEKRKEELIGLSKORLRTYSTICTRTRICVOYDQIRQPIA
VCGNKALRGKGLVTSVGNARVGAFFLNKINDTQKEFTVNEYDEGMNRLSDIQT
GQGLTREFEKGCVTSVPMKELMRRENVINPEPLMDLRFKNSYLSKVRYSKRRGA
ALLNKIQVAHNSRLRVADKVDSDGSTDSPKKGDKHKANEMWVRPTGKYSKELS
ALHMKOPIQAHGCAVMTIKESQDAHYLASGADRVTHVVEVQCELSMNEGSLTPIH
PSLCDSSGNEITVVEKKKKGSGRRRNHPIDIVHPETVFSOKPYCSLKGHLDA
ILDLSWSKQLLSSSSMDKTVRLWDIETKCLKLFAHNDYVTCIQSPVDENTFLSG
LDKARTWSQDRHVSDMLHEMTAACTPDQCALIHGICRAYDTECKLSQ
TNGIDVOSNKSQAKRKITSFQSPVNPSEVLVTADSRIILDGSEVIRKFGFRNT
CNSLASYSODGKYIICASDSQVYLKNDHFRTSLTTQSHHEHFC KDVSAAPWH
GHVRGEPPIQHSKHSKRLSTSSOPSSITSSPTKEETSATGPTTSNRNKKSGLPPH
PKRAATKSQIQPEEAGPELGSESFSKSNSEQSRRFSGESFINSRLSSWSWF
DSGGHGQTIQPTAGWGVIVTAVHGEIRSYQNFGLPRRIGRQTTLF"
complement(56145..56194)
/rpt_family="n"(GAA)n"
56686..56744
/notes="exon predicted by xgrail, quality
marginal_shadowexon"
complement(56743..56808)
/notes="exon predicted by xgrail, quality marginal"
complement(57092..57128)
/notes="exon predicted by xgrail, quality marginal"
57429..57497
/notes="exon predicted by xgrail, quality
marginal_shadowexon"
complement(57759..57834)
/notes="exon predicted by xgrail, quality excellent"
complement(join(<57887..58388,59719..59846,60078..60132,
60218..60291,60384..60527,61543..>61581))
/genes="F13M22.18"
complement(<57887..>61581)
/notes="predicted by genefinder"
/genes="F13M22.18"
complement(join(57887..58388,59719..59846,60078..60132,
60218..60291,60384..60527,61543..61581))
/genes="F13M22.18"
/notes="hypothetical protein"
/codon_start=1
/db_xref="PID:g3236250"
/translation="MPVRVVSNTPAQAFSGTGNVSNQOKEEARVNVVOIGIDLEHG
YLCGTMEALNVPADPTVITFWEGEIVDGKNYTFYCKWEATREDDMRHSSPSP
LQWKGVTDGCLTITAGFYVFCSDSGISGFIYDPNPSHHMIISSKNVL
KMBEVSKRKKTQDSDELSPKSLKCFANVACSENTNGNSEIDTEYSMSYVN
STTSMECNDEMKEESGCGDKMISFESHDIYVYQTNLEDFSEKVIENTIYLDE
QEEDDEAKGSSNAKFLVSSGRWTVNQKENRILV"
58770..58795
/rpt_family="n"(GA)n"
59165..59234
/rpt_family="AT_rich"
59249..59312
/rpt_family="AT_rich"
complement(59446..59478)
/rpt_family="AT_rich"
complement(61713..61851)
/notes="exon predicted by xgrail, quality good_shadowexon"
join(<62177..62406,62814..63035,63131..63242,63337..63399,
63504..63642,63940..63995,64072..64145,64247..64394,
64470..64652,64757..64844,64948..65058,65257..65354,
65591..65708,65788..65882,66002..66108,66213..66306)
/genes="F13M22.19"
/notes="receptor-like protein kinase"
/db_xref="PID:g3236253"
/translation="MFLKULTIFFFFFFNLIQSSSQSLNFAYNNGFNPPDLSIOGI
TIVTPNGLKLTNTTVQKTHAFYKPIRFKDSPNGTSGVTSFVFATHSQIAISLG
HGIAFYVAPNAPSLGNPNQSGICLFNLANNRNETNHFVAVELDTILSTFNDFNDHV
GDINLSKVSQSPAGYQWDEQKQNLTIISRKPQMVDYDQRTNKIDVTMAFPED
KTRPLVTVAROLDISVLLDMXVSGSSATGSVLSYEHYILGWSFGLNEKAPLASRLP
KLPREPFRKISSEFYKIGMPLISLILFISFELVLCYIVRRRRKFAELEWEKEFGNR
FRKDLIYATKGFKEKGLLGTGTGFGSVYKGVPGTKLEINAVKRVSHESRQGHKEFVNE
IVSIGHMSHRNLVPLUGYICRRRGELLVLYDIMPNGSLDKILYNTPEVTLNKQRIKI
```



```
/product="putative alcohol dehydrogenase"
/db_xref="PID:g3236259"
/translation="MANAIFFLKNTGAKFPVSLGTWQASGLVGDVAAAVKIGYR
HIDCAQIYNEKEIGAVLKLFEDRVVRKEDLFIITSKLWCTDHPDQDVPALNRTLKD
LQLEVYDLIHWPARIKKGVGIPKPNLLPVDIPSTWKAEMALYDSKARAGVSNF
STKKLADLELARVPVAVNOVECHPSWRQTKLOEFCCKSGVHLSAYSPLGSPGTWLK
SDVLKNPILNWAELKGSQAQVALRWGLQGHVLPKASTNEGRIKENFVFDWSIPD
IMFAFAIEQVRS"
complement(91371..91445)
/rpt_family="(CAAAA)n"
92640..92726
/feature="exon predicted by xgrail, quality excellent"
complement(93519..93682)
/feature="exon predicted by xgrail, quality good_shadowexon"
93640..93819
/feature="exon predicted by xgrail, quality excellent"
complement(93688..93779)
/rpt_family="(CAT)n"
BASE COUNT 29945 a 17475 c 16793 g 29632 t
ORIGIN
Query Match 7.1%; Score 70.4; DB 39; Length 93845;
Best Local Similarity 74.2%; Pred. No. 4.9e-05;
Matches 89; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 363 attcttggctccaaagattcttatacatgtagcttggtaaaagggaatttagatctgc 422
|| || || || || || || || || || || || || || || || || || || || || ||
Db 84594 ATCTTGTGGCCGCAAGATCTTATACGTGTAGCTTCTCCGGAGAGATTTAAATCGGC 84535
QY 423 tcaagctctgttgacacataatgttcatagagagatagacgacatttgagacaatc 482
||||| || || || || || || || || || || || || || || || || || || || || ||
Db 84534 TCAAGCCCTCGGTGCCATATGAACGTTCAACCAAGAGACAGACGCGCTTAAGCAAGC 84475

RESULT 6
DDU32174 5931 bp DNA INV 22-JAN-1997
LOCUS Dictyostelium discoideum spore lysis A (spIA) gene, complete cds.
DEFINITION U32174
ACCESSION 9974333
KEYWORDS
SOURCE slime mold.
ORGANISM Dictyostelium discoideum
Eukaryotes; mitochondrial eukaryotes; Dictyosteliida;
Dictyostelium.
1 (sites)
Tan, J.L. and Spudich, J.A.
Developmentally regulated protein-tyrosine kinase genes in
Dictyostelium discoideum
Mol. Cell. Biol. 10 (7), 3578-3583 (1990)
90287147
2 (bases 1 to 5931)
Nuckolls, G.H., Osherov, N., Loomis, W.F. and Spudich, J.A.
The Dictyostelium dual-specificity kinase spIA is essential for
spore differentiation
Development 122 (10), 3295-3305 (1996)
97053827
3 (bases 1 to 5931)
Nuckolls, G.H. and Spudich, J.A.
Direct Submission
Submitted (21-JUL-1995) Glen H. Nuckolls, Departments of
Biochemistry and Developmental Biology, Stanford University, B400
Beckman Center, Stanford, CA 94305, USA
Location/Qualifiers
1..5931
/organism="Dictyostelium discoideum"
/strain="JH10"
/db_xref="taxon:44689"
/chromosome="4"
861..5763
/gene="spIA"
join(861..1001,1083..1156,1224..5763)

gene
CDS
```

```
/gene="spIA"
/standard_name="spore lysis A"
/feature="exhibits tyrosine and serine/threonine
autophosphorylation activity; also called DpYK1"
/codon_start=1
/product="non-receptor tyrosine kinase"
/db_xref="PID:g974334"
/translation="MSEKFFKKLVPSNEYRWDLRKNSLTINIEDSKRSYRLPTSG
SKGIATQTPFSSTFYELFITNGDKICFGLTINDHPLEIYVPGYQSGYSGDG
KCYFGTNEGRVIGVPSFGSDVGGYSSKTLVFTKNGVYVGAOKVNLIGLPTV
GLQNGPSVINFEFGPFSYRCAPKPSQSTIKDGGSSLIIPSEDLIPKPEFEVCRWS
EKKNYHCKHVVRNRTAFPLDPSKDTIGVRATQPGEGCFEYFIIDQLDKQLSI
GLANLEPTTHVGMPSRSGYHNDGKFRWREEPGNEGESYSSYKKGDIIGCGGL
SFTSREIFFKNGMYLGFATSNVGYFVPSFAFNEPGISITGVGPPFKFSQVTLMLK
NVNSTSLVPGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SNNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SIGNSSIIQRSPPHSINGPLMLPPSSTNNNNNIYSSYNSSTAGSSTTILPTLNHP1
FGNTTNNSSSTLSVGGNNLGRHCQSLPTASTNHTLSSSLGVSFGSPSSPKTS
PRKTVNSSEDLGFVQTDQDQPPPSAMRRCGSKSIKDKDITLTIIKKTSVAMADRP
SEIGEPYGVFFFDGDSIASSCVIGCINTSTRDIFFTKNGHYLGVAFSRVSOTLPY
FSSNSSSTICYFEVYLEGHDKGSITVGLSHSTYFFIKHIGREPKSYCFSEGEKYGG
SIISFRGVVGLCVATFPGHHFENIEDLPGISPSVWTEALGDPORGSGQFNWANDVA
IWLESFNGOYRKNFRDNNISGRHLEGIHAMLKNLGLIEPYCHREDIIRLNMIO1
WNKSPSYPKIAIDSSDKIRWPSAGSGSSGGINISGGVIGSSSGSDGDIETISSSK
NIRPKSYTQKEIEDNRRTISGEEKNKIYIDNQMDPHQIGSMDSLPLDFGQGP
PDKNSSKTLSENIQIRYLOQKDEPPIAISSTNGSVSSTGGSSGFLTFPSSNSLTH
PPORDKPTQFTLHPITTSYNGITNGQPHKSPDQPLELFPRHSAFSSNNGNNNN
NNNNNNKANOQOQOQSSYQSOQOQOQHTSTSTTKWIDDPFGGWETOSLSHP
PSRPPPPPPPPQLPVSEYEIDFENELFQGTIGKGFGEVKGYWRETDVAIKLIYR
DQFKSSLSLVNEQVILSKLRHPNVVQLGACTAGGEDHHCIVETEMWGGSGSURQL
TDHFNLEQPHIRLKLALDIAKGMNLYLHGMTPEILHRDLSRNLILDHNDIPKNPV
SSRQDIKCKISDFGLSKRKEQSQMTQVCIPYMAPEVFKGDSNSEKSDVSYGMY
LFELTSTDEPQDMKPMKMAHAAVESYRPPILPTLSTSSKKWKEILLTQCWDSNPDSPF
KQIIVHLKEMEDQGVSSFSFVQTIIDTVYA"
1002..1082
/gene="spIA"
/feature="intron confirmed by sequencing RT-PCR product"
1157..1223
/gene="spIA"
/feature="intron confirmed by sequencing RT-PCR product"
5259
/gene="spIA"
/feature="reported as C in mRNA sequence, GenBank Accession
Number M33785"
/citation=[1]
/replace="C"
5310..5311
/gene="spIA"
/feature="reported as GT in mRNA sequence, GenBank Accession
Number M33785"
/citation=[1]
/replace="gt"
BASE COUNT 2247 a 1037 c 857 g 1790 t
ORIGIN
Query Match 4.9%; Score 49; DB 3; Length 5931;
Best Local Similarity 47.3%; Pred. No. 1.1;
Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY 216 caacagtaagtacttcaaaaaaagcatcatgagacagatggagtagtactgaataa 275
||||| || || || || || || || || || || || || || || || || || || || ||
Db 722 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 781
QY 276 caacaatggcgacaataacaacaataatgtttacaagctcattacagataattatgg 335
||||| || || || || || || || || || || || || || || || || || || || ||
Db 782 CAATAATAATAACAACAACAACAACAACAACAACAACAACAACAACAACAAGAAG 841
QY 336 aaatgaagatcattacttctgtgactattctcttctggcctccaagatcttatactag 395
||||| || || || || || || || || || || || || || || || || || || || ||
Db 842 AGATTCTAGTTGGTTTAAATATGAGCTTTAAATTCCTTTAAAAAGAAATTTGGTCCCAATCAA 901
```

```

Oy 396 cttttgtaaaaggaatttagatctgctcaagctcttggtgacacatgaattgtcatag 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 902 TGAATATAGATGGGATTTAAGAAATCAATCTACATTAATATTGAGATAAAG 961

Oy 456 aagagatagagccattttgagacaatcaccacccctagagatatgaatagggtattctctt 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 962 TAGATGTTCTTACAGACTACCAACATCTGTTCAAAAGGTGTAAGTATATTTCTTTTTT 1021

Oy 516 aaacctaattctt 528
    ||| ||| ||| |||
Db 1022 AAAATTTTATTTT 1034

RESULT 7
LOCUS AE000787/c 38829 bp DNA BCT 10-DEC-1997
DEFINITION Borrelia burgdorferi plasmid lp38, complete plasmid sequence.
ACCESSION AE000787
NID 92690175
KEYWORDS
SOURCE Lyme disease spirochete.
ORGANISM Borrelia burgdorferi
Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
burgdorferi group.
REFERENCE 1 (bases 1 to 38829)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kierlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi
JOURNAL Nature 390 (6660), 580-586 (1997)
MEDLINE 98065943
REFERENCE 2 (bases 1 to 38829)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
Richardson,D., Peterson,J., Kierlavage,A.R., Quackenbush,J.,
Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
Direct Submission
JOURNAL Submitted (08-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source 1..38829
    /organism="Borrelia burgdorferi"
    /db_xref="taxon:139"
    /plasmid="lp38"
    482..667
    /gene="BBJ01"
    482..667
    /gene="BBJ01"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BBJ01"
    /db_xref="PID:g2690206"
    /translation="MKYHIIVSIFVFLNACPNDSNTNQNSKELKTGRIPNKQIK
    NALLDDLKNIETASAG"
    927..1211
    /gene="BBJ02"
    927..1211
    /gene="BBJ02"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1

gene 4576..5496
CDS 4576..5496
    /gene="BBJ08"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BBJ08"
    /db_xref="PID:g2690188"
    /translation="MEMNLLINRFILNFIFFDINRGVPLLYRDKAIAASRPKIRFSA
    ILI"
    4576..5496
    /gene="BBJ08"
    4576..5496
    /gene="BBJ08"
    /note="hypothetical protein; identified by Glimmer;
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BBJ08"
    /db_xref="PID:g2690188"
    /translation="MFLYLLTLGLISCNDSLKPNKEQNNNDIKETLGSSVOENAL
    NNLYGNOEKKDFKFEELKDESLIALAKSLASTRPTTVCNIESAVLPVGHVVSLET
    ANKVSPTISIKHNOKEKKEDLSPTREKKAKRAIKADIENLRDSDGFPFELIESY
    SLAHEYTLIRNNFYDITIRNRKRLSLIRNGRNNRDKIKELTQLQNNKLKIVDELDEIM

```

```

/transl_table=11
/product="B. burgdorferi predicted coding region BBJ02"
/db_xref="PID:g2690207"
/translation="MINHLCLCKKETDLFTVSDLKKLKNSEFKLSIKTDISKMLEQ
LLDYKDYINSTDIADKLESHLTLEYKQIVKRSFQTTELKNNILSISNL"
1593..1745
/gene="BBJ03"
1593..1745
/gene="BBJ03"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ03"
/db_xref="PID:g2690184"
/translation="MKAKKHPYIYNFLNVSLTGLIGSFWRGLYWQHGGELLRSQVL
GEIYLR"
complement(2268..2381)
/gene="BBJ04"
complement(2268..2381)
/gene="BBJ04"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ04"
/db_xref="PID:g2690185"
/translation="MLLYKKTFSMIGFCFRALSENKHFMLLFSAGTIF"
complement(2765..3261)
/note="This region contains an authentic frame shift and
is not the result of a sequencing artifact: similar to
GP-2228714 percent identity: 94.44; identified by sequence
similarity; putative; transposase-like protein, authentic
frameshift"
/gene="BBJ05"
3486..3632
/gene="BBJ06"
3486..3632
/gene="BBJ06"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ06"
/db_xref="PID:g2690186"
/translation="MISSKSFIFLLSYLENYCVLPFPLIYLKNNYMQSLSOAGHYK
ELNY"
complement(4164..4307)
/gene="BBJ07"
complement(4164..4307)
/gene="BBJ07"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ07"
/db_xref="PID:g2690187"
/translation="MEMNLLINRFILNFIFFDINRGVPLLYRDKAIAASRPKIRFSA
ILI"
4576..5496
/gene="BBJ08"
4576..5496
/gene="BBJ08"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ08"
/db_xref="PID:g2690188"
/translation="MFLYLLTLGLISCNDSLKPNKEQNNNDIKETLGSSVOENAL
NNLYGNOEKKDFKFEELKDESLIALAKSLASTRPTTVCNIESAVLPVGHVVSLET
ANKVSPTISIKHNOKEKKEDLSPTREKKAKRAIKADIENLRDSDGFPFELIESY
SLAHEYTLIRNNFYDITIRNRKRLSLIRNGRNNRDKIKELTQLQNNKLKIVDELDEIM

```









ANYYELQTVSLRTNLNTHSLQFYFENNFLILKECSLKVSNILEFNKKINLKDPFFY  
NMFSIELKIEL"  
23018. .23122  
/gene="BBJ30"  
CDS  
23018. .23122  
/gene="BBJ30"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ30"  
/db\_xref="PID:g2690196"  
/translation="MWYIFSYIQDECIMKLGADKIIRKQDRYCLKT"  
23366. .24088  
/gene="BBJ31"  
CDS  
23366. .24088  
/gene="BBJ31"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ31"  
/db\_xref="PID:g2690197"  
/translation="MEKGGHNPVVGKMFESAIEIKALSKEKDNPKDMVEKASIS  
AKESDKVJUSSKRGNASIALGEVPSLVRDIRESAKKIDESLKLIIASGYDASYPLSDN  
MKIGRICLLDKILEAVNPKINDSDNDFDSCSEKIKKVVDEFNKEKDEYSCSIM  
SLNSVPKTKGKDLVKKCIINELMKDVEKYLGVKGNGVCGELLOGAKSASQGFESL  
KLLSPAASSIAEACKRLAYDIL"  
24681. .24794  
/gene="BBJ33"  
CDS  
24681. .24794  
/gene="BBJ33"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ32"  
/db\_xref="PID:g2690198"  
/translation="MLNSLSKADINIIVRRESTILLKGLDYSAISALFVLKRFWV"  
24681. .24794  
/gene="BBJ33"  
CDS  
24681. .24794  
/gene="BBJ33"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ33"  
/db\_xref="PID:g2690199"  
/translation="MLLCVNVILNNFSLVILKEQLFNKKNVNYSESSEGC"  
24681. .24794  
/gene="BBJ34"  
CDS  
24681. .24794  
/gene="BBJ34"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ34"  
/db\_xref="PID:g2690200"  
/translation="MIKGNFTFILIIVTMFVSKFYGSDDTNKKNTSLNGDTPREIDNI  
GSVILEQDNKGGDTASKVALDQVTEHANSELMNDPDDSSISRYNQENTGTKLTEE  
MDKLKAFVKITTYQGLINSIYNNKTSYNTIATYSGCANYNSICFSEGSARRSQ  
ALNDLEKKNLDEYTKLNOMLAETQDVCPKALDNAIEYKPAITIAKEADKIKKIT  
SFTIDEGNNNEERKENVDNLKKNVILSISEKTIETASVAYANAFIVSRLSAEFI  
TAAVEFKATRYANGKGDHADVVGVAIGAIADFNENRFRANFANKGAEGVDK  
MIAAEKLRATYTAVPKKNKDK"  
26858. .26962  
/gene="BBJ35"  
CDS  
26858. .26962  
/gene="BBJ35"  
/note="hypothetical protein; identified by Glimmer;  
putative"

/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ35"  
/db\_xref="PID:g2690201"  
/translation="MKVCERREEFYKINVLFKDSSIIDKENVIEYRT"  
28053. .28053  
/gene="BBJ36"  
CDS  
28053. .28053  
/gene="BBJ36"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ36"  
/db\_xref="PID:g2690202"  
/translation="MKRKSNICISLLVTLFVSCKFFGNKSASKEKETSESDDTASKI  
SKSGTAASSDQKENTSVTDGAKHTSSPYMLADALVSDTTNDRDKQENKDKLNE  
EDKKLNAFFSTTKTYOSSLSIYNNKYTYINTIDYGCSDTYRIECSFVGSEKRRQ  
ALADLEKLDKTYQLSMLKSAPVSYKKNLDDSIQAYKEAIKQAIKQAEASKIETVK  
DYATAQSAADDEKKRNIDNLKIVRDVLLIIRKTIKASRSYADAFATSSLSCEFK  
QAVREFNDAAKQYANGKGNVAVIVGTISSMPYVKFKDEFARAKMFARNVRCGDEV  
KNIRPIDKLCDVYKKVAL"  
28442. .28442  
/gene="BBJ37"  
CDS  
28442. .28442  
/gene="BBJ37"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ37"  
/db\_xref="PID:g2690203"  
/translation="MILLLLPVVLVNLGKLLKILVFAKEYANGKCKIMILMLLGA  
ISSWAYNEFE"  
28703. .28703  
/gene="BBJ38"  
CDS  
28703. .28703  
/gene="BBJ38"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ38"  
/db\_xref="PID:g2690204"  
/translation="MALLVNFESKMLKDDVPYKDNQYFRECNCVK"  
29264. .29401  
/gene="BBJ39"  
CDS  
29264. .29401  
/gene="BBJ39"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ39"  
/db\_xref="PID:g2690205"  
/translation="MKISSIVVLGMITFSVISEIISPRQFYSLOEFCSQLYNIKKLK  
K"  
29922. .29922  
/gene="BBJ40"  
CDS  
29922. .29922  
/gene="BBJ40"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BBJ40"  
/db\_xref="PID:g2690215"  
/translation="MYEHKEKLLILLFLKYKRVQHIIFLRLISLV"  
30771. .30771  
/gene="BBJ41"  
CDS  
30771. .30771  
/gene="BBJ41"  
/note="similar to GP:1553115 percent identity: 25.93;



```

Qy 769 aaggaagattagacttggaacttcgactggatgccacttaactatctatctaatctttacg 828
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17722 AAAGCACAAATTATTAATTCGGAATTAGTTTTCGAATCAAAATCATTGATTTATGCGATCA 17663

Qy 829 gcattaaagtttgtaaatgagtcagacagcttagcctaaactactatgcactttaatat 888
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17662 TTAATTAGTAGAAGACGTCACACAAAACCAAGACATTAAGTAATTTAAGTAAATTATAT 17603

Qy 889 ggctctctgtgctatattattattttttacatggtgatctagttgctatttaagat 948
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17602 TGAATACCTTCCTCTCTATTATTCATAAATAACATTTTCTATGTTTGAAGCTCATTT 17543

Qy 949 ttagtacctgttcagattaaaagaaacgaagtgtaattaaaaaaa 996
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17542 GTAAAAATTTTATCGGGTGAAAAAAATCAAAACTCAAAATTACCAGAA 17495

RESULT 8
AE000787/c 38829 bp DNA BCT 10-DEC-1997
LOCUS Borrelia burgdorferi plasmid lp38, complete plasmid sequence.
DEFINITION AE000787
ACCESSION AE000787
NID 92690175
KEYWORDS Lyme disease spirochete.
SOURCE Borrelia burgdorferi
ORGANISM Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE 1 (bases 1 to 38829)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
            Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
            Winn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
            Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
            Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
            Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
            Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
            Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
            Genomic sequence of a Lyme disease spirochaete, Borrelia
            burgdorferi
            Nature 390 (6660), 580-586 (1997)
JOURNAL 98065943
MEDLINE 2 (bases 1 to 38829)
REFERENCE Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
            Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
            Winn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
            Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
            Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
            Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
            Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
            Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
            Direct Submission
            Submitted (08-DEC-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
            Location/Qualifiers
            1. .38829
               /organism="Borrelia burgdorferi"
               /db_xref="taxon:139"
               /plasmid="lp38"
               482..667
               /gene="BBJ01"
               482..667
               /gene="BBJ01"
               /note="hypothetical protein; identified by Glimmer;
               putative"
               /codon_start=1
               /transl_table=11
               /product="B. burgdorferi predicted coding region BBJ01"
               /db_xref="PID:g2690206"
               /translation="MKYHIIVSIFVFLNACNPDSNTNQNSKELKTGRIPNKQIK
               NALLDDLKNIETASAG"
               927..1211
               /gene="BBJ02"
               927..1211
               /gene="BBJ02"

Qy 769 aaggaagattagacttggaacttcgactggatgccacttaactatctatctaatctttacg 828
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17722 AAAGCACAAATTATTAATTCGGAATTAGTTTTCGAATCAAAATCATTGATTTATGCGATCA 17663

Qy 829 gcattaaagtttgtaaatgagtcagacagcttagcctaaactactatgcactttaatat 888
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17662 TTAATTAGTAGAAGACGTCACACAAAACCAAGACATTAAGTAATTTAAGTAAATTATAT 17603

Qy 889 ggctctctgtgctatattattattttttacatggtgatctagttgctatttaagat 948
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17602 TGAATACCTTCCTCTCTATTATTCATAAATAACATTTTCTATGTTTGAAGCTCATTT 17543

Qy 949 ttagtacctgttcagattaaaagaaacgaagtgtaattaaaaaaa 996
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17542 GTAAAAATTTTATCGGGTGAAAAAAATCAAAACTCAAAATTACCAGAA 17495

/genes="BBJ02"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ02"
/db_xref="PID:g2690207"
/translation="MINHLCLKKETDLTIVSDLKKLNSFEKLSIKTDISKMLEQ
LLLDYKDYINSTKTDIAKLESHLTELYKQIVKKSFTTELKNILSISNL"
1593..1745
/gene="BBJ03"
1593..1745
/gene="BBJ03"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ03"
/db_xref="PID:g2690184"
/translation="MKAKKHPYIYNFLNVSFLTIGSFFWPRGLYWQHGLLRSQL
GEIYLR"
complement(2268..2381)
/gene="BBJ04"
complement(2268..2381)
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ04"
/db_xref="PID:g2690185"
/translation="MLLYKKTFSMIGFCFRALSEKHFLLMFLFSAGTIF"
complement(2765..3261)
/note="This region contains an authentic frame shift and
is not the result of a sequencing artifact.; similar to
GP-2228714 percent identity: 94.44; identified by sequence
similarity; putative; transposase-like protein, authentic
frameshift"
/gene="BBJ05"
3486..3632
/gene="BBJ06"
3486..3632
/gene="BBJ06"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ06"
/db_xref="PID:g2690186"
/translation="MIISSKSFIFLLSYLNYCVLPFLIYLKNNYMQSLSOAGHYK
ELNY"
complement(4164..4307)
/gene="BBJ07"
complement(4164..4307)
/gene="BBJ07"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ07"
/db_xref="PID:g2690187"
/translation="MEMNLLINRFILNFIFFDINRGVPLLYRDKAIASRPKIRFSA
ILIT"
4576..5496
/gene="BBJ08"
4576..5496
/gene="BBJ08"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ08"
/db_xref="PID:g2690188"

```

```
/translation="MFLVTLTIGLISCNLDKSLPNKEKNNNDIKETLGSSVQBNAL
NNLYGNQEKDKFKEBELKDESLLALAKSLASTPTTVGNLESVLPVGHVLSLETS
ANKVSPTISIKKKEKREDLSPSTKEEKADKAKIDENLHDSGFFELIESWY
SLKHETTLIRNNFYDVIKIRNKKTSLIKGNRRNDKIKELTQIQQNNLKIIVDELDEIM
VHIDIAEOEIRSAEFFNEAKELKEGIIKRLSESNKVASQARQALNKVEDALKSLE
ASSSKRGLAMGRRIIRIKELIENAKTVLSKS"
gene
6089. .6862
/genes="BBJ09"
CDS
6089. .6862
/genes="BBJ09"
/notes="similar to percent identity: 99.61; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="outer surface protein D (ospD)"
/db_xref="PID:g2690176"
/translation="MKKLTKILLLSFLLSLSCVHDKQELSSKSNLNKQGYLDNEG
ANSNYESKQSLTSLNOLLKOTNTSLKEAKNTONTNLSANEANKVWEAVINAVNLIS
SAADOVKATKNHDLAOMAEIDLEKIKNSDPKATFASNLAKEAYSLTKAEONMOKL
YKEQKIKISESESDYSDEAIKQAKEAVEIAKATVEAKDKLIDVENTVKTLDKIK
TETTNNTKLADIKAEAEVLQIAKNAKEIVQEVVALLNT"
7270. .7476
/genes="BBJ10"
CDS
7270. .7476
/genes="BBJ10"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ10"
/db_xref="PID:g2690189"
/translation="MGARDEFKNGIKLDEKTERARIELSTKITDVRSELSQVKGGLI
FPFYILGIFIPAVIGFLYLLQK"
7780. .7965
/genes="BBJ11"
CDS
7780. .7965
/genes="BBJ11"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ11"
/db_xref="PID:g2690190"
/translation="MFLSFCNLCTIKNIHIGSLRVRTVFNVDYAFWLSKTAEDYFP
DAKDVICIRGKKKPYI"
8633. .8725
/genes="BBJ12"
CDS
8633. .8725
/genes="BBJ12"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ12"
/db_xref="PID:g2690191"
/translation="MVLGIWKAENYVDKFDLIIYKKNELFVI"
8877. .9155
/genes="BBJ13"
CDS
8877. .9155
/genes="BBJ13"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ13"
/db_xref="PID:g2690192"
/translation="MLHINHANEAFKEISSAKDKDSSDIGKIVDVQKTHNRLHDL
RDIKTEFYQKNSFLNGVRAEKSKKKNQILFRSTSSINGKIRNANYIL"
9125. .9286
/genes="BBJ14"
CDS
9125. .9286
/genes="BBJ14"
/notes="similar to PID:g2180 percent identity: 33.33;
```

```
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="PID:g2690177"
/translation="MLHLDLATNFVIFQLSIYNLFVTFEFKSSNAEPDPLVPFDL
LLFKDSSKP"
10040. .10168
/genes="BBJ15"
CDS
10040. .10168
/genes="BBJ15"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ15"
/db_xref="PID:g2690193"
/translation="MSISWLEIIFKFIQSOTIDNVQKININKTGVEVALIKLKLNI"
10518. .11105
/genes="BBJ16"
CDS
10518. .11105
/genes="BBJ16"
/notes="similar to GB:U03641 PID:458218 percent identity:
35.19; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="PID:g2690178"
/translation="MKRAEKDKALFNFRCDCNEETNLNDDOKELANYNNLKEOL
KYNLKDIDNNKIQRMIYKELYKYGDFARNFDIKSFEVAKSOAYRYLKIYOK
VLECKSDIKYEGCFKAILDPIKAKDSLEDNHSSEGANESIPRILVRDKELYNE
CKQDTRKYFIEIKYKRDVLSELIYEKNNK"
11152. .11889
/genes="BBJ17"
CDS
11152. .11889
/genes="BBJ17"
/notes="similar to GP:2182756 percent identity: 39.43;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="PID:g2690179"
/translation="MDRKSNITITIANLKGVGKSTLSILFSYVLKDLGKKVLLIDMD
SONAUTSFRKYVFNHKNYINLLIGNVYFDCISKINDNFIIPSPHFDENDKN
LDNKNLTSFCLDKNVLGDFDYFELDTPPSFSFILKNALNTNHIIPVQPTWSIE
SLEILMKITDKSYNISIVNQFIKRNRLKAEVEDALYKRYSNYIKGKIHYVNSIKVF
IINRLPDIKSKYKEAKDVLKNILD"
11862. .12452
/genes="BBJ18"
CDS
11862. .12452
/genes="BBJ18"
/notes="similar to GB:U03641 PID:458217 percent identity:
39.36; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="PID:g2690180"
/translation="MNDLEKLARKKEIISKKEAYNNNINKGTERTAFPRIEID
NKKIYTKIFKVLKFRANKDNKLSLTFOKLNNKNYLFNLPKEDNKELGIKYG
WDKLEKPFRLQNNKSYIKKLYLEFSPKSGSIKCYVOSLTLRLKKDKESTKYKF
NUEHIKRMENVYKFSKLNKNGVIYKWIERNQIL"
12445. .13440
/genes="BBJ19"
CDS
12445. .13440
/genes="BBJ19"
/notes="similar to GP:1655797 percent identity: 27.04;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="PID:g2690181"
/translation="MALNLLNNHNKNTNNTNLLSKDSRLKKIISVINYNKSEFKKYD
ISYIRHFNSEKLELYPHHQIDLRVLNSNISKEGYKPTVIRTREDLFLIHKAI
```



```

DKKDFGTGALOFDNEAYLQGLFLLNYKILEFGIAPSFIVQNNDOYFSFNKLFNFNLSPND
FIFKLRONYLNGLIENIVLARTTIEPEWFFEFYFISNYSVLSGSMLDKESLDKF
SSPKYLSPLHYFQASNEVDLLAMVEFPSEIENKTEDIVIFELFEIYSGIFFYSTL
RQDLWSKNFKDDDDNRFLLGURYIISFENDVNDLSVFESISRNNNFIFGAGIKL
AMIYELLQTVLSLRTNLNTHSLQFYFENNFLILKECSLKVSNILEFNKKINLKDTPEY
NMFSIELKIEL"
/23018..23122
/gene="BBJ30"
CDS
/23018..23122
/notes="BBJ30"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ30"
/db_xref="PID:g2690196"
/transl_table=11
/translation="MYYIFSCYIQDECIMKLGADKIIRKORYCLKT"
23366..24088
/gene="BBJ31"
CDS
23366..24088
/gene="BBJ31"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ31"
/db_xref="PID:g2690197"
/transl_table=11
/translation="MEKGGGSHNPVVGKMFESAIEIKALSKEPNKDMVKASIS
AKESQKVLSSKGNASIALGEVPSLVRDRESAKRIDESKLIIASGYDASPLSDN
MTKIGICRILDLLEAVMPKINDSONDFDNCSEKIKKVVDFNKNKDEYSCSIM
SLNSYPTKGKDLVKKICNLEMKDVEKYLGVKNGVNGVCGELLOQAKSAGQFGESL
KLISEAASSAEACKRLAYDIL"
complement(24386..24517)
/gene="BBJ32"
CDS
complement(24386..24517)
/gene="BBJ32"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ32"
/db_xref="PID:g2690198"
/transl_table=11
/translation="MLNSNLSKADINIIVRRESTIILLKGLDYSISALFVLKRFHV"
24681..24794
/gene="BBJ33"
CDS
24681..24794
/gene="BBJ33"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ33"
/db_xref="PID:g2690199"
/transl_table=11
/translation="MLLCVNVILNNFSLVILKEQLFNKKNVNVYSESSGC"
complement(25337..26407)
/gene="BBJ34"
CDS
complement(25337..26407)
/gene="BBJ34"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ34"
/db_xref="PID:g2690200"
/transl_table=11
/translation="MIKGNFIFILVTTFVSCKFYGSDDTNKKNKNSLNGDTRIDNI
GSVILLQDGNKGDTTASKVALDOVTEHANSMLANDDPDSSISKYNOENTCKLREE
DMKLRAPFKVTITYQCILNSIYNKTRSYNTIATYSGCANYANSICGFCSPARRSQ
ALNDLEKNKDEYTKLQNMKETTQDPCPKALDNLAEYKRAITAKEAEDKIKKIT
SFTIDEGNNEKERNVDNLKNVNLISSEKTIETASVAYANAFIVISRLSAAEFI
TVAVEFKDATEKYANGKGDHVDVVVGAIGAFADNENRFERAKMFANKEGAEDVK
MIAAIEKLRATYAVKPKNKDK"
26858..26962
/gene="BBJ35"

```

---

```

26858..26962
/gene="BBJ35"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ35"
/db_xref="PID:g2690201"
/transl_table=11
/translation="MKVCRREFFYKINVLFKDSSIIDKENYIEYRT"
complement(26995..28053)
/gene="BBJ36"
CDS
complement(26995..28053)
/gene="BBJ36"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ36"
/db_xref="PID:g2690202"
/transl_table=11
/translation="MKRKNICISLLVTILFVSCKFFGNKSASKEEETSFSDTASKI
SKSGTAASSDKQKNTSDVTGDAKHTSSPYMLADALIVSDTTNRDRDKQENKDLNE
EDKRLNAFFSTTKYTOSSLDISYNNKYTYNTIDYSGDTRYIECFYVGPSEKRRQ
ALADLEKLDEKYTQLSMLKSAPSYKKNLDDSIQYKEAIIQAIESKIEYVK
DYATAQSAADDEKKRINDLKIYRVQVLLIIKKTIKASRSYADAFATATSSLSCEFFK
QAVKENDAAKOYANGKGNVAVNVIVGTISSMPYVYKDEFAKAKMFARNYRGDEV
KMLRAIDKLCDYKKVAL"
complement(28278..28442)
/gene="BBJ37"
CDS
complement(28278..28442)
/gene="BBJ37"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ37"
/db_xref="PID:g2690203"
/transl_table=11
/translation="MILLLLPVYLVVNLGKLLKILVFAAKEVANGKIMILMLLGA
ISSMAYNEFE"
complement(28605..28703)
/gene="BBJ38"
CDS
complement(28605..28703)
/gene="BBJ38"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ38"
/db_xref="PID:g2690204"
/transl_table=11
/translation="MALLVNFSKMLKDDVPGYKDNQQYFRECCKV"
complement(29264..29401)
/gene="BBJ39"
CDS
complement(29264..29401)
/gene="BBJ39"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ39"
/db_xref="PID:g2690205"
/transl_table=11
/translation="MKISIVVLGMITFVSIVISIIIPROFYSLQEFCSQLYNIKKLK
K"
29827..29922
/gene="BBJ40"
CDS
29827..29922
/gene="BBJ40"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BBJ40"
/db_xref="PID:g2690215"
/transl_table=11
/translation="MYEHKKEKLLILFLKYKVQHIFLFLRLSLV"
complement(29905..30771)

```





[illegible]



```
/note="var(40.0)"
/codon_start=1
/transl_table=3
/db_xref="PID:g343949"
/translation="MKKLLNIIILSIKNTNNNNIIINNTLDSLNNKKLLKKNILLD
INNKINIKRILNNINPAGNVVHRIGPAGNINNKLOHNNINNNXTOIYNYK
NIEIINTNDKLKLLYKIITLKNINIKIISITNOHSNKLNIKFVYVNDI
NNNNNNYIINIKLINIINNINNLNLSYYNKKVTIEPIKLSYIYLNDF
SKYISLDINGILNIYORIILNNIPKUNDHNSINYINNINNNNNNNIINL
LNNNNNNNNNNNNYIGNNIYNNITIDNIPIDILYKLVGXSKFKGRILS
NNNGRTSTNLLNGNNKYLXNNNNYKLYIPSHNNLYNNSINKNGYNIKVK
LNFI"
BASE COUNT      605 a   57 c   67 g   471 t
ORIGIN            About 125 bp upstream of hinf1 site.

Query Match      4.7%; Score 46.8; DB 38; Length 1200;
Best Local Similarity 45.2%; Pred. No. 4.2;
Matches 171; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

Qy 203 atagaaacagttccaacagtaagtaacttcaaaacaaagcatcgtggaagacagatgg 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 ATATTAAATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 850

Qy 263 agtactgaataacaacatgagcaataacaacaaataatgttacaagctcattac 322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 851 ATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 910

Qy 323 gagataattgaaatgaagatcatttacttgggtggactctctcttggccctccaagat 382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 911 ATAATATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 970

Qy 383 ctatcatagtgactttgtaaaaggaatttagatctgctcaagctcttgggtggacaca 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971 ATTTAGTTGGTGATCTATTAAATTTAAAGGTAGATTAAGTAATAATAATGTTAGAACTA 1030

Qy 443 tgaatgttcataagagagacccatttggagacaatcaccacctgagagataataata 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1031 GTACACTTAATTATAATGGTACTTTTAAATAATAATAATAATAATAATAATAATAATA 1090

Qy 503 ggtattctcttcaacaccttaacttgaacaaacctaaactttacccttagtcataacc 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1091 ATAAATAATAATAATAATAATAATAATATCCCTCTCAATCAATAATTAATAATAATCTAATA 1150

Qy 563 ctagtttttcaagaaat 580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1151 TTAATAAAAAATGGTAAAT 1168

RESULT 13
PETZFB2      PETZFB2      1083 bp      mRNA      PLN      10-MAR-1995
LOCUS        Petunia mRNA for zinc-finger DNA binding protein.
DEFINITION   D26084
ACCESSION    D26084
NID          9439488
KEYWORDS     zinc-finger DNA binding protein.
SOURCE       Petunia hybrida (strain:Titan parade) cDNA to mRNA.
ORGANISM     Petunia x hybrida
              Eukaryotes; mitochondrial eukaryotes; Viridiplantae;
              Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
              plants; Magnoliophyta; Magnoliopsida; Solanaceae; Solanales;
              Solanaceae; Petunia.
REFERENCE    1 (sites)
AUTHORS      Takatsuji, H., Nakamura, N. and Katsumoto, Y.
TITLE        A new family of zinc finger proteins in petunia: structure, DNA
              sequence recognition, and floral organ-specific expression
JOURNAL      Plant Cell 6 (7), 947-958 (1994)
MEDLINE      94348284
REFERENCE    2 (bases 1 to 1083)
AUTHORS      Takatsuji, H.
JOURNAL      Unpublished (1995)
COMMENT      Submitted (06-Dec-1993) to DBJ by:
              Hiroshi Takatsuji
              Department of Applied Physiology
              National Institute of Agrobiological Resources
              2-1-2 Kannondai, Tsukuba
              Ibaraki 305
              Japan
              Phone: 0298-38-8383
              Fax: 0298-38-7417
              Location/Qualifiers
                1. .1083
                /organism="Petunia x hybrida"
                /strain="Titan parade"
                /db_xref="taxon:4102"
                109..933
                /codon_start=1
                /product="zinc-finger DNA binding protein"
                /db_xref="PID:d1005618"
                /db_xref="PID:g439489"
                /translation="MALEALKSPTAATPSLPPTYEDHVDNNLDSWVGKSKRPRIE
                TPSEEYLALCLIMLARSNGTTPGSTDTITITISKEPEKKNRELTTPVHOETEQSY
                KCSVCDKFSYQALGGHAKSHRKITTIATLDDNNNNPTTSNHSNYYNNISAL
                NPSGRSHVCSYCHKAFTQALGGHKKRRHYEGKLGNNNNHRRDGGHSGSVVTTSDGG
                ASHTLRDFDLNMLPPSPPELQGLSICDLKSOIQIEQEVESPMPLKKPRLFSMD"
                328 a   216 c   198 g   341 t
                BASE COUNT
                ORIGIN

Query Match      4.6%; Score 46.2; DB 8; Length 1083;
Best Local Similarity 69.2%; Pred. No. 5.6;
Matches 63; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 381 atcttatcatgtgcttctttaaaggaatttagatctgctcaagctcttgggtggaca 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 ATCTTACAAGGTAGCTGTGTGACAGCTTTTATGTTCTTATCAAGCTCTTGTGTGACA 467

Qy 441 catgaatttcatagaagatagagccatt 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 CAACCAAGTCATAGGAAAAATTCAACTATT 498

RESULT 14
PETZFB2      PETZFB2      1083 bp      mRNA      PLN      10-MAR-1995
LOCUS        Petunia mRNA for zinc-finger DNA binding protein.
DEFINITION   D26084
ACCESSION    D26084
NID          9439488
KEYWORDS     zinc-finger DNA binding protein.
SOURCE       Petunia hybrida (strain:Titan parade) cDNA to mRNA.
ORGANISM     Petunia x hybrida
              Eukaryotes; mitochondrial eukaryotes; Viridiplantae;
              Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
              plants; Magnoliophyta; Magnoliopsida; Solanaceae; Solanales;
              Solanaceae; Petunia.
REFERENCE    1 (sites)
AUTHORS      Takatsuji, H., Nakamura, N. and Katsumoto, Y.
TITLE        A new family of zinc finger proteins in petunia: structure, DNA
              sequence recognition, and floral organ-specific expression
JOURNAL      Plant Cell 6 (7), 947-958 (1994)
MEDLINE      94348284
REFERENCE    2 (bases 1 to 1083)
AUTHORS      Takatsuji, H.
JOURNAL      Unpublished (1995)
COMMENT      Submitted (06-Dec-1993) to DBJ by:
              Hiroshi Takatsuji
              Department of Applied Physiology
              National Institute of Agrobiological Resources
              2-1-2 Kannondai, Tsukuba
              Ibaraki 305
              Japan
              Phone: 0298-38-8383
              Fax: 0298-38-7417
              Location/Qualifiers
                1. .1083
                /organism="Petunia x hybrida"
```

[illegible]

```
/gene="MJECLO3"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO3"
/db_xref="PID:g1522646"
/translacion="MRQCKVQVGGFGGWIKCDLELFFDVMRGEIEBELVEFIKSLSES
LLKSSSDSRKYNLALISTLTKLTPWTSRLEYADDEHNIDVYPDDKNNKWIICNL
CLPDE"
1934..3253
/gene="MJECLO4"
1934..3253
/gene="MJECLO4"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO4"
/db_xref="PID:g1500652"
/translacion="MFVDRREELKALNEKLDSSNNFEFIVYGRRRICKTKIALKSVEN
REHYIYLAVEDGNDLKHFRYASKVEPIEYAKEDWEAYFNFLDKIILIDFFPNLIE
NPVLSLQFRIVDIHLKNTKTLILGSSISMGEKVLSPIDYGRKTGLVKIKPLK
FKHKEFFPAINEELVEIYFADGIPYLEKVKLPFWDLKREIKRVDSPFLREYDF
LMKYFEPEPTTKYLKALIAFGNHTLGEIKNLGFGHSDLTPLYKLNIEVEFIERQTP
ITESVKKGRIYIKDNFIAPYFPIFNLSAIEGIFDIEEKADYNOLVGFVFEKV
AKEPLILNKMLPKFLKIGRWHRGEEIDLJALNDOKKALFVSVKMDLKDRDV
KKIYRDIYRKSCLVGLDDYKYYAIVGKKIESKNGDCLFLDLEDS"
3265..3453
/gene="MJECLO5"
3265..3453
/gene="MJECLO5"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO5"
/db_xref="PID:g1500645"
/translacion="MEIALLIEGIIINDKKVAERFLDKLESSQGMWKEIRERAE
RAKQLEGEIEMAKTKL"
complement(3784..4134)
/gene="MJECLO6"
complement(3784..4134)
/gene="MJECLO6"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO6"
/db_xref="PID:g1500653"
/translacion="MSVELRTLFRLIAVLEHSEFFKVLFCACERHFESGCKCGPMEM
CNIALACAMKEDPTLVLRKWRVFTYLEEVGIITRKLEAPANRPRYIKLSENWMEA
LRTAIDXEYELIIS"
complement(4538..5805)
/gene="MJECLO7"
complement(4538..5805)
/gene="MJECLO7"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO7"
/db_xref="PID:g1522647"
/translacion="MKMFIIISKNHLDIYNFVDFEKNFKNISSIGNIINDEVVMNLPK
GVEVCGDGRCKVFCGIVYGLSSAIGKNIKMGPELGYLPFFEKEDRVRLM
MTLEYRLATVSKNVDLILLDGLSALIMPLPSGDTNPILTYVPDLAEDLGWAFIKS
LDNFWEVLENDKNDIYDNTLLAIKIFQKDSITYSEVDEIREELFAANILNLSLEIA
CWGYFYIELLNLRLLEDYCTFIANKFENSITTEKLKNNINVDILLDATLLNOI
FRGRGYTTLKDCYCNKRKRSHINKICDVGDFYKFLVIRENPFEMIPKTYVYRFAE
SSPILAEVPRTNKSTEEVLSLIPYSKLGYPRLYKDAHNKAKISKKEFKKQILFMI
KYISEKNRDFSLEFQSGREVLG"
complement(5829..7391)
/gene="MJECLO8"
complement(5829..7391)
/gene="MJECLO8"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO8"
/db_xref="PID:g1522648"
/translacion="MMVINSVVVGTVVASKNWNFEFVIEQVIDIKKKGFEVITKNT
HGDIILSKITIVSNALIGKDSADSELAIRGVIYSEMLANSSKFLASAKILGVI
SRFASGIESNVYIPNPQNYLTAKDILAFSGSIEVGYLKVRSSTVKYKNAKEIC
SRFASVAMTGAQSNITAVLOELFEKDKRNNIVDPHGVYKRNTHILPAKILN
PILVPEHLAKLIGIGENSSVQKSFVYAAITVKYECKENKKOISGLLEYLKKITEEKLIV
ECADTKANSDKKRIYIEYVDGTCRKTIVKDDMSINRVIEKLRFWIKNNKILGE
NEGWFIDSKINVLPLQKIDDEEATIVGEFLKRVLKERIKSVHDIIVEIKALEKPT
LVITIEHAHFAAKNLKDRSGYWINRIAKGRFGVGLGVSRQPKELNPVLVSOMNTK
IILRIVEPTDQKYLISESSNVGDELQDLQPLSTGEAIVVVGSSLPALVKIKYIDGV
LGGEDGLKNLKI"
7780..8106
/gene="MJECLO9"
7780..8106
/gene="MJECLO9"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO9"
/db_xref="PID:g1522649"
/translacion="MNYKFMWNSDEGFIVEEQHKEEIPFEDMIVNTVEOLKLNQYTR
YTPEELTPKTNLEAYNKLEILKKHTMDIENNVRKRIYHCIAENNNKYEIAFYIVI
SKWNE"
8128..8787
/gene="MJECLO10"
8128..8787
/gene="MJECLO10"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO10"
/db_xref="PID:g1522650"
/translacion="MHILFRCCQMIETIHKNFRGIRELKNLENGOINIAKNNAS
KSSILEALFLSAKEGFLIKILREILLRWGYGKSIYDLFYKNSKELEVSVKFL
NODFANLTANSQSPANKNAVLELSDKNSWGRFDSHLHPDYISLITSAEATOS
NEFEISLTLIKFGYIESIYSQAYETQVLODAIRLLREAYPEVKSLSPLQYNNKWIH
V"
8818..9162
/gene="MJECLO11"
8818..9162
/gene="MJECLO11"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO11"
/db_xref="PID:g1522651"
/translacion="MGGFKSALLIALLTSLKNGYLLIDSAEAFHHPSSLEITSQML
TKSVKNNVOVFLTTHSLELIDFLEHASREGIEGRILYMRDGENLISSWESFENVR
EMRETLGIDLRG"
9165..9890
/gene="MJECLO12"
9165..9890
/gene="MJECLO12"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO12"
/db_xref="PID:g1522652"
/translacion="MRILLLEGTDVAFFIPILKLYGFSEICSDGLIIRAEKMGDISK
PICLENDVXLIYVHSGGSKQKHALTAMLTAKMGYLSNIKILGIARDIDQEHAVN
WTSLIKNAGFEVHSGDKFLIIEIDLNKIAYLVANLYDEDDDFNIPSELPDEKREAVT
DMAKEISIIKFKNSLESLSNDAERRLKPRKDITHVLAIAKNFDSGMSGLYRKFIIEQ
INNNKNVFLITICILPCLTIF"
10678..12486
/gene="MJECLO13"
10678..12486
/gene="MJECLO13"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJECLO13"
/db_xref="PID:g1522653
```

```

/translating="MINCYNFENKAYEDQKKNVKNISLIPNLSHANNEVVVKISM
KKLANLGLVSTLSSITGYKVRDWIELKEELQKSGENKCCNIVLCDDYPEEFE
21498. .22022
/genes="MJECL19"
/notes="identified by GeneMark; putative"
/codon_start=1
/translating=11
/product="M. jannaschii predicted coding region MJECL19"
/db_xref="PID:g1522657"
/translating="MESKEDDDKIKIISMTPKSEKSDKIVMOIVVEEINSPDDRYK
FVLLKVAELICKNGLCNAFLILDKIPDSYKSSALYKMAIDLYRNKKEHDLQIAE
KIPDDYKSKSEVLLKVVELLCEGKDEAINIAEKIPDNYYKSEALFKIETLSNKGYY
DRAVEIAEKIPDNF"
22850. .23293
/genes="MJECL20"
22850. .23293
/genes="MJECL20"
/notes="identified by GeneMark; putative"
/codon_start=1
/translating=11
/product="M. jannaschii predicted coding region MJECL20"
/db_xref="PID:g1522658"
/translating="MGKCRHNGEVSTFGVRPASFPVFPENLMDRIGGFVILDELWLRR
WEEIIEYPMKIPITLYPIEDYGIPTVEDMDLIVDFIKYHVSNGREVVSIGGHGRTG
TVLAWAGLNGVENPIEYVRECYCECAVETESQEEFVMEYLAKRL"
complement(23295. .24593)
/genes="MJECL21"
complement(23295. .24593)
/genes="MJECL21"
/notes="identified by GeneMark; putative"
/codon_start=1
/translating=11
/product="M. jannaschii predicted coding region MJECL21"
/db_xref="PID:g1522659"
/translating="MARGVMVYQFQFINRNCPLARDIILKFFPIVIERKNIEKV
EDILKVRANKYSGETANKDCAMKFEEDVGRHEALKEALFAIIFPINCNDLUSE
SLOLFVFNAYMRPFSSYSSIRVIRHGFAGVCDRNFNHNNTWISNCYNL
RMLGDFYSDTLDPYINPSCNITNKGGLYIISNTQITASLYVLLNLRN
GNIDYANLIRIRGVGDIAFLRDISIHNMLYGCROLLYKPIORVVKVIEE
LVKCKLNVMDINEVLEKFDYTCRCIQKLRNKAIDSNDFKTLVPYKIFLYEYSK
KCKLDORYVEMGMFNSKAKYVNRRLCKYIOEACKDLIKRLKAINNTNIRYIGN
NDFENIRKIIIVSEDELILKLGELIYDELLDILFF"
complement(24851. .25090)
/genes="MJECL22"
complement(24851. .25090)
/genes="MJECL22"
/notes="identified by GeneMark; putative"
/codon_start=1
/translating=11
/product="M. jannaschii predicted coding region MJECL22"
/db_xref="PID:g1500667"
/translating="MLELRSLSIRGGIAVLECGRRCSAINLKTGDKIWEFTWEDIE
SISIKDRNVMKNCGRHHYIDLKTGRKIRELIL"
complement(25133. .27616)
/genes="MJECL23"
complement(25133. .27616)
/genes="MJECL23"
/notes="identified by GeneMark; putative"
/codon_start=1
/translating=11
/product="M. jannaschii predicted coding region MJECL23"
/db_xref="PID:g1522660"
/translating="VVMRVFVENKINGVNTROEDNLNVSNYKEIKYDKFKNVFSKN
ROIFNYVDKLIELMKLNDNLESLDEKMLCKLSDIIEFKAGSVWDLSDIKDIK
DNITILGCSNHLFALDIKTGNKIWEKVEHNVDSLFKDNVYMLEYRGHGVCLDVMY
GDKIMESKVERGMWGLADNIVILGDGQIYIADVRITGGKLEFEAEWVWVLSIK
DDKILRCEDEYGCYFYVLDIKTEKILEFGGHNVSLLISGDVITLADMMGCYVA
LNTNLYSKTORVPOLNTMKVILDTLLKSLNLEWDELPIQITNKSLDITIS
KTSINEEDILFKEPIKRGDVKINLFNPKVKGKLDIDIVVEFEDEFNIRKE
RTEVITPKFGDNVDMRPEKIDKILKEIDNLTNSYKEIYSREKNIENRAVF
NYMMKILVNSNDELAINVKGKILDILGIRVDELLWEPRAEGVRLLSIKGDIY
ILGCVSGHYVAIDIKTGKWEFKAEEDTVWGLSDIDIVVLCGNGIFESIVMLKNSKI
LLEGVGHVADINTGIRWSKIKHDSVRSIKDDIVVLCGKGYILALDINAGNMLWE
FRAKSGKSRNLSIKNDILLFGCDNYLYALDIDITGTRELWRFRKAEGEVKSLSIKNDVNL

```

```
LGCRGYYVLLDINTGKMERFKVGSVLRSLSIKDDIIVILGCNRCVYALDINAGENL
WAKTDCGVNLSKNDVAVLGDNYLAIYDINTGEIMFKTESAVSALDLSIKDNIVI
SGCKRGHVAYLWKNLKNYSIIQIKQVL"
28069..28848
/gene="MJEC124"
28069..28848
/gene="MJEC124"
/notes="similar to SP:P37522 percent identity: 35.7;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC125"
/db_xref="PID:g1522661"
/translation="MVVISANOKGGVKTIALNLSFTLAEGYDPLVIDLDPQENL
SFGILNKKLDYADKNIGILLNSNVYKKKIEISLIKNDKLLIPLSHLQSAVERKL
VNAYARKKUNIIQENKENDYIIIDNAPSLGLFLINSVADSYIIIPCPEYSFSA
GVOLMDVEIEKESNPKLVGLGFIFNKYSKQIPKRLPOLKQLYPNIPVIGI
PRTITVEKAREGKVPKFDANNPASVAFSELAEWIENVK"
28838..29170
/gene="MJEC125"
28838..29170
/gene="MJEC125"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC125"
/db_xref="PID:g1522662"
/translation="MSNDUNALKKKEISSGVVQITSLPKPEOEKMSKTLRKN
TTHEKIEMFGKVGSGEVVDGVAVLWALPKLPEQKRVVLAEDRFEAFADR
LGIEIKKE"
complement(29175..30206)
/gene="MJEC126"
complement(29175..30206)
/gene="MJEC126"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC126"
/db_xref="PID:g1522663"
/translation="MGIMFYDREKELNYTCOLEPNSILFVYGPKSSGKSTVMR
RVKLESNIVFYNYLRATPTREDFLRVFEXSDKKYLLKLENLNGVCKFGIE
ENFDNLCNDVAFKINSINAVVEGKPVLIIDELQKLNIFYNGKSLNELFN
LFVSLTMEHLCHVICLTSOTLFDIETRNSTLKNASEYVLIDWLARKGTNRNLKEE
FSSEINVALDLSLPEISQLNNKLGISVQTIQWLNVEKDKLYLIDTIDDEE
EGYNVLSRKDKIKISIDREVKRKEMKYLFLENIILFYDINGIIRKITSIIENHA
IKEIINAMQ"
complement(30568..31056)
/gene="MJEC127"
complement(30568..31056)
/gene="MJEC127"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC127"
/db_xref="PID:g1522664"
/translation="MDILGEIVNDEIVEKLEQKNDIGYKPFCDYTNADAKVVRK
HVSKHLEEELKLEKSKSNKKGTQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
VLFFCHNHKVRHLANGEVLEGKCCCKDPYTLVDVGKGDVIVNKAIVYKVPDLDE
KL"
complement(31531..35349)
/gene="MJEC128"
complement(31531..35349)
/gene="MJEC128"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC128"
/db_xref="PID:g1522665"
/translation="MNNLSNSKSLRYWDQVKAIFDKYQINVSIGYINISDILES
DLXHLIAQLFLVLKHEEFVSVLGETHLEKILKLVQKQFYGLIEIRFLGFKGKQFF
IDISDLHVNLIKTEIEIKLIKQVEYIKKQKRLDGGVYGVIAPRKRGDGAAT
AWSYIAFDFOVEEKNMPTSEIMEKLIKLYLSKVEKDIPLHKVAFGGGIRFLY
PERPILIEILLMIADLGDADIAMDLARVDRLVGTYNKEKYGSPRCVITIAQLT
```

```
DEEDIEKLLTPILYKFGIENYKEFEVLENEKREELNKRGLNNTTTINRLIKKP
NLNKTLODKNRIYKWLTYIQDKLTKLKGKWEIKLLSYGLDGYKNGNTPDLWS
LPDDKNDPDCSIYINEGNANVNDPDMRIFALAGLWMIKEFKDKLIEFKLHNI
NPKSTYRTVIREIMNELLRETIKIEAGYLPKEAIKAYQUSIEQGVVFKADTGR
GKTYTLTIGTREIKVEFKKHAJAVAPYKIQVLQVAGLHADGVVPMYVEDGKLDK
NTIHYLTITQDOVENMLDQDVTYKGIQVANEEDILLAIIDEADLVIOKEFRKR
AIVGVKCDRAGGCVLLATPELINLNNVPYIEVEFKDEKKLFERCSHIAKNIIGE
FCEYILLRPOGWIKNAVLDNMKNENIKYTLGYGKPKYIVYTRVETVGDKASK
MINEEKVPEGLILATRVISEGVNKNHVDLWALYCKSATIRFIACRNGGSEL
IVTAPFEREEPMIIDYNAMIEMFKENYKLLKEYLETIDIEVLKELDKNNKVLISOI
ONAIYYDEEKKOMWDETAIHVLSLEAHITRDYKLLKEYLEKTYGEFAIKTKE
LKESKDLKFLKYNLEYLSEKVSAKHIIIAFKYGVNEIKNALENNYKPKCDRDEVT
POLIKHSHRINSINVLDQVNIPEVIVNYMLNKGDDKKSLNDENTIKKCLEE
INERYKIVINEENEDRKKKYHIDKLEEHKDKSLARIIEELLILKTFCLCPPSK
WGIQIRIIRAVNIVVGEDDRLRRKLSLVAKVILKVRDFAIEROKFRKIRELIE
WKECGFTLTLDVRRLLRAIFNCVIRGVKLGENAVVEIKELNEFKTYLTIION
INSNVEKEKVEIKKIEENGGEILEMELYELIVERKFEVEEYKALEKLLKGLVIY
EPKPGLLRIT"
complement(36217..36420)
/gene="MJEC129"
complement(36217..36420)
/gene="MJEC129"
/notes="similar to GP:U08838_1 percent identity: 64.2;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="archaeal histone"
/db_xref="PID:g1500647"
/translation="MTELPVAPFERILKKVGAERVSRAAAEYLAEEFAIEATAKEA
VDLAKHAKRTVKVEDIKLALKK"
complement(37148..37609)
/gene="MJEC130"
complement(37148..37609)
/gene="MJEC130"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC130"
/db_xref="PID:g1522666"
/translation="MSPNVFIDSSVMVGLFIDGDEKAHLLSKLINDGMLCINPIVF
SELMFKVTHIALEDIRGVLDLKNLNNYSWYEEVREKIDKMKNNVLLKIDTWE
VKLAPEICKKYNLLNDGAIITATCKYIRINKLATDSDFEKVDFEIIKE"
complement(37596..37802)
/gene="MJEC131"
complement(37596..37802)
/gene="MJEC131"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC131"
/db_xref="PID:g1522643"
/translation="MIYSEIIEIYEDVGLKPLKPKLKIKGKRLKIKIVNDDVEEFL
KSMIICKDKIDYKLEAYYESF"
complement(38825..40141)
/gene="MJEC132"
complement(38825..40141)
/gene="MJEC132"
/notes="identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC132"
/db_xref="PID:g1522667"
/translation="MGKELDNRTLDFFYFKYKKKTSTEPSYKDIIETALISENLE
DYQSQHTKLINILNLYPKYPPHLYDLGYRPPKSGTLGVRIUDGYITKNGNPA
EITPDVILVNDNDTHVLIPECKSKSIKKEQLEKYLKKNLHVINKYISVTKNPL
YKSDYSYMSFDLTHPVLKDFEDIOILKVSPEGEISLERNGENQKNIFFITWGEK
EKPSYDILICDAENDREGSKLILRELVSMAISGHDEPEVGEVSKKLIILPGEYS
LNSIKMDSLVKEFTVGTIMDALYSTSISIIDYKDYKNIVSKRIBETLFLFSNR
NLKGYLKKVSKRWIKVICKRMARKFESCKEEICAIGKSWQTTLD"
complement(40122..41348)
/gene="MJEC133"
complement(40122..41348)
```



```
/gene="MJEC133"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC133"
/db_xref="PID:g1522668"
/translation="MIMDTIFVGSAPDITNRPOLYEWLERNMERSYKDILDDTKLE
PONVVKTYILESNIHPKLSRIKTKIPLEDFEYMLRKADYVGLSTRKHDTPE
YLDNVEERANKLSISSNGESVELIKIISQSSDIEQFEKLIDQFOLLNDTKIKKI
ENTIKKLEHLMRTKSKIKIVSDDDKFVLEIYVVGKFTVKGTGSIERKHEIVD
KTIENVEKIEILIEDSLDYSDDGDRIPLIYEFEKEIEDLKNFVEGLISYKPEPKI
WGLAKVEDMYIISGVDLHNGDRFSLEVPWMRLYLPKSGCGNTALRLLSNIQOTY
DSEITLVEKSYGERIR"
complement(42228. .43115)
/gene="MJEC134"
complement(42228. .43115)
/gene="MJEC134"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC134"
/db_xref="PID:g1500677"
/translation="KKHMPGLTEKKLYVYDERVPYSWIPHNVAKHISKELEKYDFE
VIDASLQRLVLEGSKDNTKEXVYVFAQVDPDLVNDPNFTSLLRRFLNAGQII
VMGDSPOQTVGRADNQKLPVPMQNVLSMNPYITTKRISLTSGIVLSLPLMWGT
KPHNKTTPPGVSGVSINPLAVADNVQYAHAFIISYKQASGFVRIYDPPINKBELVT
KQIRIGILGAVTRDVTPTIWNKIEILSEEVNKIDKFDTKFKNTLASEIDTFKTTINEI
LKELEILVIOKKCENKE"
complement(43112. .45004)
/gene="MJEC135"
complement(43112. .45004)
/gene="MJEC135"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC135"
/db_xref="PID:g1522669"
/translation="MKIEQVKIENFRSIVDLTVFERGKVTVLIGKSSGKSTLNVIRI
IFENDLSTENKLNITITTEQNLKDOIIRRLWFYVONYPIKLAVLNFDEGEGDKLR
EVCGLSENVKAYIEVSIESNNQIILKQNVIIFFGYSETSKIAKEALKMKETHSK
VYGHNIESTGSLVPCVIVESGDTINKDVFNIDILKNKKCYNTFPEIKWKNIP
IKGLFAHPWVSELINVTIKNTDFARLMLYDCUREIKGSVDTYEQDPTBILTKSY
EGMSNLVFGSGDQIDIGILASISKGGKHIFLIEPEIHLHPAYIRLARVLDELDA
NLNVOLIIVTSPDPFIATLKDREFVIGVRKDSIKINFEFYTELLATRIFFYPPFRGSR
LSESALMDLVGVPGVFLSDVALIVGNSDRIILOQYIDLLQKLNLPKFSYIITPY
AORSLEKMLVKPEKDYGIKVFILADNDEGLNVAIKDMGLRENFEEVFLIEKKDILG
YIKPDEFLALDELRECLKDYEEVLQDBEEKSIIDEKIFGACKNEGKNEKDLHL
KLSGRLFTVLKNVDEEFLITLGDWDSGRSRIERYLKCKIAEKIVSRUTEVEDIKKLIILN
IDSELG"
complement(45391. .45834)
/gene="MJEC136"
complement(45391. .45834)
/gene="MJEC136"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC136"
/db_xref="PID:g1500679"
/translation="MDLMAYWLCITNEDNWVKIEKKTWGVAERYKNTINKYKVGDKL
IIFYEQSRDKNPPTRIGRVYEVSEYDKSSKIFKPPRNPNEKFPYKVKLUKEIKVE
EPPINFKELIPKLFITNKKRWSGHLMGAMREIPEEDYKLIIGN"
46155. .46868
/gene="MJEC137"
46155. .46868
/gene="MJEC137"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC137"
/db_xref="PID:g1522670"
/translation="MNNFSKNLEECFORYGYNFILLAGYIAKYPKYEETSNNTKFL
NEYHNEVSNITTCILHVYKLNLAEGVRDFENLDLLYLDNVEIKKVKSMVEPNVISDAIT
```

```
VTLVNDYNNLIDAAARNVKKGDKESYTKFYIALGYVFDVILIKENYAVKVSYKLVGIL
ISKTFGYKVIYKGGSTALKPIESCTHWRISGEINSMPSKLYNNSDKLTNISIKINTS
KLYNKSINTIRSKINEVSS"
complement(47194. .47973)
/gene="MJEC138"
complement(47194. .47973)
/gene="MJEC138"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC138"
/db_xref="PID:g1522671"
/translation="MANMQSLTNIEVQRFHCDEWEYFKFEDDEFKNLWNEIEKTLGRD
FINLTSAYFQNLVYMLGKEFKLVLVDNTNIFISQVLSYVTKGELPWILDFINNPFIE
LYAPOLIVDELKNKIENLPKCKKNNDENKASKAIKIANIISNIIKINDKKSN
WSKIAYNLIGHRDVKDIPFVLALSLDTHGIIITRDKDFDKOKIINKWVGVEKVVLTE
LSQGSFSCIMNVTAFLAFKICTSIITIIILIVTSIIKKTN"
complement(48326. .49381)
/gene="MJEC139"
complement(48326. .49381)
/gene="MJEC139"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC139"
/db_xref="PID:g1522672"
/translation="MPILEHIQLNDELKDTKHLAEILKYKNNSIKTIINVLEWEDS
NIRCYEKNVYVYFTFFIIVGLWAIPEVWLWCEQVFCISPTIIHICCLYVTDINP
ILFLPLCGVGTFLHMLWATFETLSKCDSKILKLEGLFTLSLWISLSLTILTKY
AICRDYAKLTSAILHLNLIKHYFLYPHVAVAKIDYDIYDOKLPYIKIDWMLK
LGKEKVIYTPVDIYNSKLFVEKYIKNNELKSEISDDILRKEEDVKRLEQTKNAE
QYNKKVEPIPKSLPIENYDEITHYSIVRSVIRSKYFNKFLTNKNVSNIEITKDEGK
FAVNVYIEIPNSIPNSK"
complement(49453. .52581)
/gene="MJEC140"
complement(49453. .52581)
/gene="MJEC140"
/notes="similar to EGAD:HI1285 percent identity: 36.1;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="type 1 restriction enzyme"
/db_xref="PID:g1522673"
/translation="MPISANKYFLLVGVFIGDKMKKEAKLNEDYVVENAAIQRKLN
GYSYKHGSELTPENRESYRDAILKNRFIAKININPWLTEALAKVYKVTVDINP
DFNMRGKITFYMILINGVKLEPKENGKKTREVKLIDFENINKNEFLVANOQEVEYVE
NGRPRPDLVYFINGIPAIPEFSPKSNQAKAFNDHKTAKMDIPOLYQAQILVY
SDGLETKYGSPISDWDRFFVWEGVESDDVEIVDNYGNTMYKTKGPNYTSLOLLM
GLFKKHLIEFLEDFIHDKKKIITATYFYFVVKKAVDRTIKSVLYGETPDDRIGIV
WHAQGTGKSIITMLFYAKKALKOKELNYPLLVFLDRLEDEQLNVFSVSFEARAL
STALQELIKTTPGGIIPATIQKFGKSKDEHPFLDTRDNNIIIIADEAHSYGTGLA
ONLRKAIPNASFLAFTAPIDYKDRSTFLVFGDYISAYPIDAKRHGVVVPYIYEARL
VELHTNEFDLEDETSERVANDPETKESIKVEFKLEKIMLTEDYLSKYSDIIEH
FNKRLQADGRAMVVTISRKVAVELYKMWITQPNAPKAVYVMSKNKSDPEDFPHIR
TKLELNAKEFKDPESDLKMWIVVDMWLTGFDVPCLHTMYFLAPMKHNSLAQAIARY
NRVFKDPGLIYDYGIADDLKSLSKSYSEARKDMDIKVIVTEENKRYKWTYSY
FKNTYKDWKMLSSDELSLTVKAYORVAKDDNTKKEFVRNVIALKLYLLARPHPT
IGKDLDFEFEMIKKMIYKVKYKRIEISQDLENDIQSLISKSISAKELVDFEMLKK
IKPELSVLSDFSELKIAIKYKDYVRDVLILANDDIRVMAKPIRERKSERLNEV
IEKTRKYVITTAEMIEELVNAKEIRKAAEGKELGLTEELAFYDLSLTPNPLTD
KKRVEKIAKEARTMWSGIYKARDWKKNKLOSIRAKLUKIILMEKIGIKDYSLINKISD
DLFEYAKNIYA"
complement(52610. .53908)
/gene="MJEC141"
complement(52610. .53908)
/gene="MJEC141"
/notes="identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJEC141"
/db_xref="PID:g1522674"
/translation="MLFMVKRFRWETFETKDETIGIKPDMWDVKKIKIDICEVAGGSPST
KIKEYWGGDIPWITPKDLANLEYIYISGRGNITEKAVKECSLRIFFPKGTILLTSRAP
```



```

IGYVAIAKNPLTINQGRFNLIIPKDCVVSELYLYLFTKTMSEYLYLKDISSGTFPELKG
STLKEVYIPSPDEQKIATVLSYFDLLIENKKKQNEIKELIAELSELFNWFTDFEPF
KNEBVFYDELDEIPKGEWKRGLDILKVESGNSNAPQREIYFENAKIPIFVRVKHLVK
GVCLESDFDNELALDQYKMKLYNEKSIIFOKSGESLKEARVNIIVPKFTAVNHLAVI
DWSMLNEKHFIYICLLRFLKEIYVSGKTTLPYLKSIDIENKYLIIIPQPILOKQFHS
LVQPLFKIINNQKIMVLRKIDRALLPKLVFVGLRVEEL"
complement(54268..55944)
/gene="MJEC142"
complement(54268..55944)
/gene="MJEC142"
/note="similar to EGAD:HI1287 percent identity: 39.7;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="type I restriction enzyme ECOR124/3 I M protein"
/db_xref="PID:g1522675"

```

```

QY   210  cagfttcaacagctagtgactctcaaaaacaacaaagcattcggaagcacagatgggagtactt 269
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   45161 CATTTCCACTAAATATAACATATAAAAAATCAAAACAAAAAACAACAAAATAAAAAATAAG 45220

QY   270  gaataacaacaatggcgacaataacaacaacaataatgttcagaagctcattaccgagataa 329
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   45221 TACAAACAAAAAATACAATCAAAATTAATAAACAACTAAAAATAAAAAACTTGTCAAATAA 45280

QY   330  ttatggaaatga 341
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   45281 GTAAATTAAATAA 45292

Search completed: July 15, 1999, 13:56:00
Job time: 3922 sec
```

Search completed: July 15, 1999, 13:56:00  
Job time: 3922 sec

gene	CDS	20177 a	9030 c	7435 g	21763 t	2 others																																																																																																										
complement(56184..57338)	/gene="MJECCL43"	/product="M. jannaschii predicted coding region MJECCL43"	/db_xref="PID:g1522676"	/translation="MNIDIEDFEKKPIYIQRILNRSCIEFSKMLKKYDKNMIGIQ ISERMEFPEERONKNEPKITLPEYRSQFWKYLALITLEYLRILKNGNNDIVL LDDPAPVLPYVQKSELEKLSKEYQLYNTHLSLNEEELDRVLIVYLDKRN TKIKRPSNQGDIIPIRALGVDRKILFKENLSKILFVEGISDKTILEGLGLETK NLKNWYTHPSLQGDIEDNEIKVRRVYSLCSNFEETKYVFLDGDQKKKEGFERDKTS KILFLGDENOEIDLDPKLYLDQVLYETRYKRIETHOLEKFKVKEIVEKLRSKSKII EELNPERFLNLGDSKVDVATIKRKYENPELANKFKEIIDCLNGKII"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNTLSKEENTITLLMPDFLE"	complement(57338..58333)	/gene="MJECCL44"	/gene="MJECCL44"	/note="identified by GeneMark; putative"	/codon_start=1	/transl_table=11	/product="M. jannaschii predicted coding region MJECCL44"	/db_xref="PID:g1522677"	/translation="MFTKWRDGMTKTLTKVTRNFRSIRKVAHNNIGDIAVFVGA NESGSKILAEIVKNDGIDPLDGGDIDVEMLKENKEYSKKPIIEAYFEIVNKDDEIT KISNIKEILAVIKNDINEDINFDMLNKKIOFLAFKQVADGLFTYIYVDDGLEDT KEFYDCKPNVTSKTPDLLENLYIEVVKVKKQNTPENQISNANVIRGNFNFNH YKRLDISIRLVNFEFEYDECKRIENIVKSPISVTSVISIPGRNIQLNPNYNIETILT KFNISIVTIFNLAKPREYVIDEEMELGAKVINKTWSNT

[illegible]

THIS PAGE BLANK (USPTO)

Result No.	Score	Query			DB	ID	Description
		Match	Length	DB			
1	45.6	4.6	58407	1	V21210	Methanococcus jann	
2	40.8	4.1	198	1	T17202	DNA-spanner oligon	
3	40.8	4.1	198	1	T17203	DNA-spanner oligon	
4	40.6	4.1	1690	1	Q92524	Nicotiana glauca ar	
5	39.2	3.9	144	1	Q12515	CSP-2 peptide from	
6	39.2	3.9	198	1	T17204	DNA-spanner oligon	
7	39.2	3.9	198	1	T17205	DNA-spanner oligon	
8	37.6	3.8	120	1	Q12516	CSP-2 peptide from	
9	37.2	3.7	2384	1	Q68909	Human c-myc far up	
10	37.2	3.7	2381	1	Q68910	Human c-myc far up	
11	37.2	3.7	2860	1	T92160	Human deleted in p	
12	36.6	3.7	8220	1	Q83529	P. falciparum Proj3	
13	36.6	3.7	8220	1	T72897	Plasmodium Proj3 g	
14	36.6	3.7	19124	1	T72882	Plasmodium var-7 g	
15	36	3.6	110000	1	V21209_08	Continuation (9 of	
16	36	3.6	64976	1	V21209_16	Continuation (17 of	
17	35.8	3.6	3157	1	Q13316	Duffy receptor gen	
18	35.8	3.6	731	1	Q31693	RsaI restriction f	
19	35.8	3.6	4486	1	Q35118	Encodes E2 protein	
20	35.8	3.6	3831	1	V52424	Streptococcus pneu	
21	35.6	3.6	117	1	Q12514	CSP-2 peptide from	
22	35.6	3.6	40352	1	V03032	MAGE-B cluster DNA	
23	35.4	3.6	9636	1	Q67190	P. falciparum tran	
24	35.4	3.6	7400	1	T09340	Tobacco mosaici vir	
25	35.4	3.6	1942	1	T33896	Carnation ACC synt	
26	35.2	3.5	5852	1	Q11710	Dictyostelium plas	
27	35.2	3.5	154	1	T25999	Human gene signatu	
28	35	3.5	3214	1	T62360	Human origin of re	
29	35	3.5	3214	1	T73287	Human origin of re	
30	35	3.5	2160	1	T85328	Truncated murine T	
31	34.8	3.5	1016	1	N93636	Sequence of alpha-	
32	34.8	3.5	8920	1	Q62924	Carbamoyl-phosphat	
33	34.8	3.5	1856	1	T63233	Human ubiquitin co	
34	34.8	3.5	1137	1	T63234	Human ubiquitin co	
35	34.8	3.5	12987	1	T91324	Arabidopsis thalia	
36	34.8	3.5	4061	1	V44867	Clone CT797_3 codi	
37	34.6	3.5	217	1	T20147	Human gene signatu	
38	34.4	3.5	454	1	T65069	Canine genomic mic	
39	34.2	3.4	4500	1	Q29684	CCV-6 spike gene.	
40	34.2	3.4	4359	1	Q50817	Canine coronavirus	
41	34.2	3.4	4359	1	Q52447	Canine coronavirus	
42	34.2	3.4	2681	1	T30870	Engineered 95 kb p	
43	34	3.4	19124	1	T72882	Plasmodium var-7 g	







I0 Q68910 standard; cDNA; 2381 BP.  
 AC Q68910;  
 DE 13-APR-1995 (first entry)  
 DE Human c-myc far upstream element (FUSE) binding protein (FBP)  
 DE variant lacking Ser 97.  
 KW DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;  
 KW promoter P1; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 27..1961 /\*tag= a  
 FT misc\_difference 470 /\*tag= b  
 FT /label= A,G  
 FT  
 PN W09419465-A.  
 PD 01-SEP-1994.  
 PF 22-FEB-1994; U01782.  
 PR 22-FEB-1993; US-021608.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Avigan MI, Duncan RC, Levens DL;  
 DR WPI; 94-294330/36.  
 DR P-PSDB; R58814.  
 DR New DNA-binding regulator of c-myc expression and its cDNA - used  
 PT to develop prods. for diagnosis and therapy of disease states  
 PT such as tumour formation  
 PS Claim 8; Page 57-59; 94pp; English.  
 CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far  
 CC upstream element) binding protein (FBP) are synonymous. A FUSE which  
 CC is required for maximal transcription of c-myc binds a factor (DROME  
 CC or FBP). The activator cis-element is approx. 1500 bp 5' of the  
 CC human c-myc promoter P1. A full length FBP cDNA sequence was  
 CC assembled from overlapping clones obtd. from cDNA libraries. Source  
 CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,  
 CC BJAB, and PMA/PHA stimulated pooled human peripheral blood  
 CC lymphocytes (PBLs). A composite cDNA is given in Q68909 and the  
 CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in R58813.  
 CC Three clones from a BJAB cDNA library and three clones from an  
 CC activated human PBL cDNA library contain the sequence in Q68909.  
 CC Three clones from the same BJAB library and two clones from the  
 CC activated lymphocyte library are lacking bps 316,317 and 318  
 CC (see Q68910). The mRNA lacking these three bps would encode a  
 CC variant protein lacking Ser 97 (see R58814).  
 SQ Sequence 2381 BP; 769 A; 488 C; 564 G; 559 T;

Query Match 3.7%; Score 37.2; DB 1; Length 2381;  
 Best Local Similarity 52.6%; Pred. No. 2.7;  
 Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 844 aattgagtcgacagcttagtcaaaactactatgcactttaataatggtcttctgtctat 903  
 Db 2207 AAATGATTGCAATTATGTGCTTGTGATATAAATGATGATGATGATGATGAT 2266  
 Qy 904 atttatttatttaccatgctgtctatcattgcttgcatttgaagatttagtaccctgtcag 963  
 Db 2267 ACTTTCACCTTCCAAATGCCCTTTGCTTACAAATAAATGATGATGATGATGATGAT 2326  
 Qy 964 attaaagaaacacgaagtttaattcaaaaaaaa 997  
 Db 2327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2360  
 RESULT 11  
 T92160/c  
 ID T92160 standard; cDNA; 2680 BP.  
 AC T92160;  
 DT 04-MAR-1998 (first entry)  
 DE Human deleted in pancreatic cancer locus 4 - DPC4 - encoding cDNA.  
 DE DPC4; pancreatic cancer; deleted; locus 4; diagnosis; human;  
 KW tumour suppressor gene; proliferative disease; bile duct; bladder;  
 KW colorectal; cancer; Crohn's disease; colitis; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers

FT CDS 129..1787  
 FT /\*tag= a  
 FT /product= DPC4  
 FT  
 PN W09726271-A1.  
 PD 24-JUL-1997.  
 PF 17-JAN-1997; U00827.  
 PR 19-JAN-1996; US-588821.  
 PA (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 PI Hahn SA, Kern SE;  
 DR WPI; 97-385290/35.  
 DR P-PSDB; W14287.  
 DR Deleted in Pancreatic Cancer locus 4 polypeptide - and related  
 PT nucleic acids, used in diagnosis and treatment of proliferative  
 PT diseases, e.g. cancer of pancreas or other organs  
 PS Claim 4; Page 65-67; 104pp; English.  
 CC The present sequence encodes the pure DPC4 (deleted in pancreatic  
 CC cancer, locus 4) polypeptide. DPC4 is a tumour suppressor gene.  
 CC Detection of truncated DPC4 protein, or of homozygous deletions or  
 CC intragenic mutations in the encoding nucleic acid, is used to diagnose  
 CC (in vivo or in vitro) proliferative diseases, especially pancreatic  
 CC carcinoma, bile duct, bladder or colorectal cancer, Crohn's disease,  
 CC colitis-associated neoplasia or chronic ulcerative colitis. These  
 CC conditions, where associated with a homozygous deletion, can be treated  
 CC by administering an agent that: (a) modulates DPC4 expression,  
 CC specifically a sense DPC4 sequence (particularly in the form of a  
 CC vector, i.e. by gene therapy), but also an antisense sequence where  
 CC DPC4 protein is over expressed or (b) mimics the activity of DPC4. DPC4  
 CC nucleic acid is also used as hybridisation probes for detecting  
 CC presence/absence of human chromosome 18q21.1 fragments. When a  
 CC homozygous deletion is detected in this region, an agent can be  
 CC administered that accumulates within, or kills, only cells which  
 CC contain such a deletion. This agent exploits the absence of an enzyme  
 CC (or other protein) encoded by a neighbouring gene and lost by the  
 CC deletion, i.e. it has a highly selective action.  
 SQ Sequence 2680 BP; 775 A; 548 C; 567 G; 790 T;

Query Match 3.7%; Score 37.2; DB 1; Length 2680;  
 Best Local Similarity 50.6%; Pred. No. 2.8;  
 Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 115 aatgctgaacataattattataacttactatttaagctagatttctcttctgtctagtag 174  
 Db 2329 AAGATTACAAATACATTATTAACTTCACTTTTAAACACAGAGTAAAGTACTCTGTGTA 2270  
 Qy 175 gttgactggagaaaatggcagcgatgataagaacagtttcaacagtaagtacttcaaa 234  
 Db 2269 AATTAACCTACCCACATGGCAGACATATTTTGAAGTATGTTCAACAATAACAACCTGTAC 2210  
 Qy 235 aacaaaagcatcatggcaagacagatggagtacttgaataaacaacatggcgacaata 292  
 Db 2209 ACAAATAACATAGGTGAATTAACCTTTTGAGAACCTCTCGCTTTAAATAACAGCAATA 2152  
 RESULT 12  
 Q83529  
 ID Q83529 standard; DNA; 8220 BP.  
 AC Q83529;  
 DT 22-SEP-1995 (first entry)  
 DE P. falciparum Proj3 gene.  
 KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;  
 KW vaccine; ss.  
 OS Plasmodium falciparum.  
 PN W09507353-A.  
 PD 16-MAR-1995.  
 PF 07-SEP-1994; U10230.  
 PR 10-SEP-1993; US-119677.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;  
 PI Wellens TE;  
 DR WPI; 95-123427/16.  
 DR P-PSDB; R70236.  
 PT New erythrocyte binding domain polypeptide(s) - isolated from

CC genes having homology with conserved regions of DABP and SAPP. The							
CC compositions are used for the treatment and prevention of malaria. They							
CC are also used in the preparation of vaccines for inducing a protective							
CC immune response in a mammal to Plasmodium merozoites (especially							
CC Plasmodium falciparum or Plasmodium vivax).							
CC	Plasmodium falciparum or Plasmodium vivax)						
SQ	Sequence 8220 BP;	3194 A;	1249 C;	1777 G;	2000 T;		
Query Match                      3.7%;    Score 36.6;    DB 1;   Length 8220;							
Best Local Similarity    55.0%;    Pred No. 5, 1;							
Matches    72;   Conservative    0;   Mismatches    59;   Indels    0;   Gaps    0;							
Oy	867	aactactatgcaccttaataatggttccttcttgctcatattttattttacatggcgtg	926				
Db	8030	AAGGATATATATATGGGTACCTTGTATGTTTATATATATATATATATATATATATATGT	8089				
Oy	927	atctaggttgccattttaagatttagtcacctgtcagattaaaagaacgaagttaa	986				
Db	8090	ATTTATATTAATAAAGAATAATAAAACAATTTTAAAAATGAAAAGAAAAATGA	8149				
Oy	987	ttaaaaaaa	997				
Db	8150	ATATAAAAAAA	8160				
RESULT	14						
T72882							
ID	T72882 standard;	CDNA;	19124	BP.			
AC	T72882;						
DT	12-SEP-1997	(first entry)					
DE	Plasmodium var-7 gene.						
KW	DBL gene family; SAPP; sialic acid binding protein; vaccine; therapy;						
KW	Duffy binding like gene; Duffy antigen binding protein; erythrocyte;						
KW	DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;						
KW	Plasmodium; ss.						
OS	Plasmodium vivax.						
OS	Plasmodium falciparum.						
FH	Key	Location/Qualifiers					
FT	exon	7317..15139					
FT		/tag= a					
FT		/number= 1					
FT	intron	15140..16205					
FT		/tag= b					
FT		/number= 1					
FT	exon	16206..17552					
FT		/tag= c					
FT		/number= 2					
FT		/note= "no stop codon given"					
PN	WO9640766-A2.						
PD	19-DEC-1996.						
PF	07-JUN-1996; UO9508.						
PR	07-JUN-1995; US-487826.						
PA	(USHS ) US DEPT HEALTH & HUMAN SERVICES.						
PI	Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;						
PI	Wellens TE;						
DR	WPI: 97-052231/05.						
DR	P-PSDB: W22475.						
PT	New malaria vaccines - contains cysteine-rich DBL family protein						
PT	binding domains homologous domains of the Duffy and sialic acid						
PT	binding proteins						
PS	Claim 4; Page 56-61; 96pp; English.						
CC	This sequence represents the var-7 gene of plasmodium. Var-7 belongs to						
CC	the Duffy binding like (DBL) family of genes which have homology to the						
CC	Duffy antigen binding protein (DABP) and sialic acid binding protein						
CC	(SAPP) conserved regions (see T72889 and T72888 respectively). The var						
CC	family of genes modulate cytoadherence and antigenic variation of						
CC	Plasmodium infected erythrocytes. SAPP and the Duffy antigen binding						
CC	protein (DABP) are soluble proteins that appear in the culture						
CC	supernatant after infected erythrocytes release merozoites. DABP and SAPP						
CC	mediate the binding of merozoites and schizonts to the erythrocyte						
CC	surface. These proteins are necessary for erythrocyte invasion by the						
CC	parasite. This sequence can be used in the compositions of the invention.						
CC	The compositions are for the treatment and prevention of malaria. and						





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 11:12:38 ; Search time 377.02 seconds  
(without alignments)  
4033.866 Million cell updates/sec

Title: US-09-156-580-1

Perfect score: 997

Sequence: 1 ccacgtgccattttttctct.....aaagttaattaaaaaaa 997

Scoring table: IDENTITY\_NUC

Searched: 2002476 seqs, 762712212 residues

Database :

EST:\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: gb\_est6:\*  
16: gb\_est7:\*  
17: gb\_est8:\*  
18: gb\_est9:\*  
19: gb\_est10:\*  
20: gb\_est11:\*  
21: gb\_est12:\*  
22: gb\_est13:\*  
23: gb\_est22:\*  
24: gb\_est14:\*  
25: gb\_est15:\*  
26: gb\_est16:\*  
27: gb\_est17:\*  
28: gb\_est18:\*  
29: gb\_est19:\*  
30: gb\_est20:\*  
31: gb\_est21:\*  
32: em\_est10:\*  
33: em\_est11:\*  
34: em\_est12:\*  
35: em\_est13:\*  
36: em\_est14:\*  
37: em\_est15:\*  
38: em\_est16:\*  
39: em\_est17:\*  
40: em\_est18:\*  
41: em\_est19:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	45.4	4.6	243	28	C92842 Dictyosteli
2	42.4	4.3	303	31	AU034213 Dictyoste

3	42.2	4.2	482	28	C91001 Dictyosteli
c 4	42	4.2	605	31	AI227944 EST224639
5	41.8	4.2	419	31	AU033391 Dictyoste
c 6	41.4	4.2	358	27	AA634656 ab26b05.r
7	41.4	4.2	591	31	AU033325 Dictyoste
c 8	40.8	4.1	404	21	AA225651 nc08b08.r
9	40.8	4.1	291	26	AA810593 ob64b04.s
c 10	40.6	4.1	554	19	C23753 Dictyosteli
c 11	40.6	4.1	274	28	C91401 Dictyosteli
12	40.4	4.1	294	31	AU034848 Dictyoste
c 13	39.8	4.0	560	23	AI199031 qi139h04.x
14	39.8	4.0	574	28	C91220 Dictyosteli
15	39.4	4.0	546	17	C22974 Dictyosteli
16	39.4	4.0	312	23	AI163084 A031p65u
17	39.4	4.0	397	31	AU034056 Dictyoste
18	39.2	3.9	472	12	N39467 Dictyoste
19	39.2	3.9	389	15	S5C5A01 Dictyoste
20	39.2	3.9	226	28	C90268 Dictyosteli
21	39.2	3.9	552	28	C91403 Dictyosteli
22	39.2	3.9	480	28	C93684 Dictyosteli
23	39.2	3.9	605	31	AU033779 Dictyoste
c 24	39	3.9	402	11	R99317 YQ71h03.s1
25	39	3.9	249	28	C94333 Dictyosteli
26	39	3.9	246	31	AU033454 Dictyoste
27	39	3.9	617	31	AU033655 Dictyoste
28	38.8	3.9	198	31	AU034673 Dictyoste
29	38.6	3.9	466	20	C25660 Dictyosteli
30	38.4	3.9	400	25	C84210 Dictyosteli
31	38.4	3.9	408	31	AU033476 Dictyoste
32	38.2	3.8	355	15	N98120 2212C3 czap
c 33	38.2	3.8	250	21	AA278587 zs77g03.f
34	38.2	3.8	440	28	C89946 Dictyosteli
35	38.2	3.8	497	28	C94312 Dictyosteli
c 36	38.2	3.8	529	30	AI089028 oy61g05.s
37	38.2	3.8	592	31	AI146745 qb92d08.x
38	38.2	3.8	526	31	AU033912 Dictyoste
c 39	38	3.8	395	25	AA682058 vull3c07.s
40	38	3.8	660	28	AA800394 EST189891
41	38	3.8	643	28	C90989 Dictyosteli
42	38	3.8	627	28	C91439 Dictyosteli
c 43	37.8	3.8	533	11	H17378 Ym38d12.r1
44	37.8	3.8	455	24	AA602349 no89h11.s
c 45	37.8	3.8	592	28	C89805 Dictyosteli

#### ALIGNMENTS

RESULT 1  
C92842/c

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

C92842 243 bp mRNA EST 03-JUN-1998  
Dictyostellium discoideum slug cDNA, clone SSF465, mRNA sequence.

C92842  
g3074718  
EST; EST(expressed sequence tag).

Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA,  
clone.lib:SS.

Dictyostellium discoideum  
Eukaryota; Dictyosteliida; Dictyostellium.

1 (bases 1 to 243)  
Urushihara,H.

Developmental cDNA in Dictyostellium discoideum  
Published Only in Database (1998) In press

2 (bases 1 to 243)  
Urushihara,H.

Direct Submission  
Submitted (05-APR-1998) to the DDBJ/EMBL/GenBank databases. Hideko

Urushihara, University of Tsukuba, Institute of Biological  
Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan

(E-mail:gxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,  
Fax:+81-0298-53-6614)

PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.  
Location/Qualifiers

COMMENT  
FEATURES

```
source
1. 243
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SS"
/dev_stage="slug"
BASE COUNT 117 a 8 c 17 g 101 t
ORIGIN

Query Match 4.6%; Score 45.4; DB 28; Length 243;
Best Local Similarity 52.4%; Pred. No. 0.38; Mismatches 0; Indels 0; Gaps 0;
Matches 100; Conservative 0;

Qy 807 ttaactatctcaattttacggcattgaaggtttgaattgagtcgacagcttagtcaa 866
Db 243 TTTTNTTTTTTTTTTCTCTATTTTTTTTTTTTTTTTTTTTTCATAATTTCTTCAA 184

Qy 867 aactctatgcactttaataatggtctcttgctgtatatatttattttacatggtcgt 926
Db 183 TTCCCCAAACACTGTTTTAAAAATATATTTTTTTGTATTAATAATTTTAAAAAATATTACAGT 124

Qy 927 actcagggttcatttaagatttagtcacctgtcagatttaaaagaaagaaagttaaa 986
Db 123 ACCTATCTATTTCATTATTTATTTTGTAAATTTAAAGAAAAAATAAAGAAAAAAGAAAAA 64

Qy 987 ttaaaaaaaa 997
Db 63 TAAAAAATAAA 53

RESULT 2
AU034213 303 bp mRNA EST 28-OCT-1998
LOCUS Dictyostelium discoideum slug cDNA, clone SLC264, mRNA sequence.
ACCESSION AU034213
NID 93799637
KEYWORDS EST.
SOURCE Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA, clone_lib:SL.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 303)
AUTHORS Eukaryota; Dictyostellida; Dictyostelium.
TITLE Urushihara,H.
JOURNAL Developmental cDNA in Dictyostelium discoideum
REFERENCE 2 (bases 1 to 303)
AUTHORS Urushihara,H.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1998) to the DBJ/EMBL/GenBank databases. Hideko Urushihara, University of Tsukuba, Institute of Biological Sciences; 1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664, Fax:+81-0298-53-6614)
PROJECT = Dictyostelium discoideum cDNA project in Japan'.
COMMENTS Location/Qualifiers
FEATURES
source
1..303
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SL"
/dev_stage="slug"
BASE COUNT 172 a 67 c 12 g 52 t
ORIGIN

Query Match 4.3%; Score 42.4; DB 31; Length 303;
Best Local Similarity 58.9%; Pred. No. 1.9; Mismatches 0; Indels 0; Gaps 0;
Matches 73; Conservative 0;

Qy 198 catgtagagaacagtttcaacagtaagcttctcaaaaacaaagcatcgtgcaagaca 257
Db 54 CAACAACAACCACTACTACAACAACCAACCACTACAACAACAACAACAACAACAACA 113

source
4
AI227944/c
LOCUS AI227944 605 bp mRNA EST 30-OCT-1998
DEFINITION EST224639 Normalized rat brain, Bento Soares Rattus sp. cDNA clone RBRCN90 3' end similar to similar to guanine nucleotide dissociation stimulator a ras-related GTPase, mRNA sequence.
ACCESSION AI227944
NID 93811831
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 605)
AUTHORS Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
```

```
Qy 258 gatggagtacttgaataacaacaatggcgacaataacaacaacaataatgttacaagtc 317
Db 114 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATAATTATACCAA 173

Qy 318 atta 321
Db 174 ACTA 177

RESULT 3
C91001 482 bp mRNA EST 15-APR-1998
LOCUS Dictyostelium discoideum slug cDNA, clone SSJ551, mRNA sequence.
ACCESSION C91001
NID 93060367
KEYWORDS EST: EST(expressed sequence tag).
SOURCE Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA, clone_lib:SS.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 482)
AUTHORS Eukaryota; Dictyostellida; Dictyostelium.
TITLE Urushihara,H.
JOURNAL Developmental cDNA in Dictyostelium discoideum
REFERENCE 2 (bases 1 to 482)
AUTHORS Urushihara,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1998) to the DBJ/EMBL/GenBank databases. Hideko Urushihara, University of Tsukuba, Institute of Biological Sciences; 1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664, Fax:+81-0298-53-6614)
PROJECT = Dictyostelium discoideum cDNA project in Japan'.
COMMENTS Location/Qualifiers
FEATURES
source
1..482
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SS"
/dev_stage="slug"
BASE COUNT 232 a 98 c 51 g 101 t
ORIGIN

Query Match 4.2%; Score 42.2; DB 28; Length 482;
Best Local Similarity 61.3%; Pred. No. 2.2; Mismatches 0; Indels 0; Gaps 0;
Matches 68; Conservative 0;

Qy 198 catgtagagaacagtttcaacagtaagcttctcaaaaacaaagcatcgtgcaagaca 257
Db 368 CAGCAACAGCAACAATAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 427

Qy 258 gatggagtacttgaataacaacaatggcgacaataacaacaacaataatgt 308
Db 428 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATAT 478

RESULT 4
AI227944/c
LOCUS AI227944 605 bp mRNA EST 30-OCT-1998
DEFINITION EST224639 Normalized rat brain, Bento Soares Rattus sp. cDNA clone RBRCN90 3' end similar to similar to guanine nucleotide dissociation stimulator a ras-related GTPase, mRNA sequence.
ACCESSION AI227944
NID 93811831
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 605)
AUTHORS Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
```



Db	317	GATGGAAGAATTCAACCAATCTACCCCAACACAAAGCAAAAGCGGAGATGAGTG	258
Qy	262	gagtacttgaaataacaacaatggcagacaataaacaacaataatgtttacaagctcatta	321
Db	257	GGGCAACGGGATGAAGCTGTAGCACACTTACAGGTAGCATAGTAGACACGCGCACAT	198
Qy	322	cgagataaattatgggaatg	340
Db	197	TCATATGATCAGGATCTG	179
RESULT 7			
LOCUS	AU033325	591 bp	mRNA
DEFINITION	Dictyostellium discoideum siug cDNA, clone SLA616, mRNA sequence.	EST	28-OCT-1998
ACCESSION	AU033325		
NID	g3798749		
KEYWORDS	EST.		
SOURCE	Dictyostellium discoideum (strain:AX4) siug cDNA to mRNA, clone_lib:SL.		
ORGANISM	Eukaryota; Dictyostellida; Dictyostellium.		
REFERENCE	1 (bases 1 to 591)		
AUTHORS	Urushihara,H.		
TITLE	Developmental cDNA in Dictyostellium discoideum		
JOURNAL	Published Only in DataBase (1998) In press		
REFERENCE	2 (bases 1 to 591)		
AUTHORS	Urushihara,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko Urushihara, University of Tsukuba, Institute of Biological Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664, Fax:+81-0298-53-6614)		
COMMENT	PROJECT ~ 'Dictyostellium discoideum cDNA project in Japan'.		
FEATURES	Location/Qualifiers		
source	1..591		
	/organism="Dictyostellium discoideum"		
	/strain="AX4"		
	/db_xref="taxon:44689"		
	/clone_lib="SL"		
	/dev_stage="siug"		
BASE COUNT	234 a 115 c 77 g 162 t	3 others	
ORIGIN			
Query Match	4.2%	Score 41.4; DB 31; Length 591;	
Best Local Similarity	55.1%;	Pred. No. 3.5;	
Matches	81; Conservative	0; Mismatches 66; Indels 0; Gaps 0;	
Qy	198	catggatagaacaggtttcaacagtaagtacttcaaaacaaaagcatcatgccaagaca	257
Db	375	CAA	434
Qy	258	gatggagtacttgaaataacaacaatggcgacataaacaacaacaataatgtttacaagctc	317
Db	435	CAACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACC	494
Qy	318	attacgagataattatggaatgaaga	344
Db	495	ACAACCTCAACCTCATGTACATGTAGA	521
RESULT 8			
LOCUS	AA225651/c	404 bp	mRNA
DEFINITION	nc08b08.f1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007511, mrna sequence.	EST	20-AUG-1997
ACCESSION	AA225651		
NID	g1846568		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		

```

DEFINITION  ob64b04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336111,
mRNA sequence.
ACCESSION  AA810593
NID        92880204
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 291)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 926 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 242.
            Location/Qualifiers
            source
            1. .291
            /organism="Homo sapiens"
            /note="Vector: p7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was prepared from human tonsillar cells enriched for
            germinal center B cells by flow sorting (CD20+, IgD+),
            provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
            (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
            primed with a Not I - oligo(dT) primer
            [5'-TGTTACCATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-
            3']. Double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not I
            and Eco RI sites of the modified p7T3 vector. Library
            went through one round of normalization, and was
            constructed by Bento Soares and M. Fatima Bonaldo."
            /db_xref="taxon:9606"
            /clone="IMAGE:1336111"
            /clone_lib="NCI_CGAP_GCB1"
            /tissue_type="germinal center B cell"
            /lab_host="DH10B"

BASE COUNT  97 a 45 c 36 g 113 t
ORIGIN

Query Match 4.1%; Score 40.8; DB 26; Length 291;
Best Local Similarity 49.5%; Pred. No. 4.5;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 786 tgaacttcgactgggacacactaatatctattttacggcattaaagttgttaa 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 TGCATTTTGTGGTATGTTATTATTCAATAAAGATTATTATTAGTGCTGAAC 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 846 ttgagtcagacgcttagtcaaaactactatgcacttaataatgctcttgcctat 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 TGGACCTGTGTGTAAGCCCGAGCAAGCAATCTAGGTAGGGTTTAATCCCGAGTAAAT 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 906 ttattttattacatggctgctatcttgccatttaagatttagtacctgtcagat 965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 TGCCATATTGCACATGCTTATGAAGTTTGAATGTTAAATGAATGTTATTCACTTTA 39
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 966 taaagaaaaacgaagttaataataaaaaaaa 997

```

```

||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 7

RESULT 10
LOCUS    C23753 554 bp mRNA EST 22-APR-1997
DEFINITION Dictyostellium discoideum gamete cDNA, clone FC-AF03, mRNA sequence.
ACCESSION  C23753
NID        92073265
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE     Dictyostellium discoideum (strain:KAX3) Sexually competent Gamete
            cDNA to mRNA, clone lib:FC.
ORGANISM   Dictyostellium discoideum
            Eukaryotae; Mitochondrial eukaryotes; Dictyostellida;
            Dictyostellium.
            1 (bases 1 to 554)
            Urushihara,H.
            Direct Submission
            Submitted (04-APR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Tennohdai, Tsukuba, Ibaraki 305, Japan
            (E-mail:d402huesakura.cc.tsukuba.ac.jp. Tel:0298-53-4664,
            Fax:0298-53-6614)
            2 (sites)
            Kobayashi,A., Shimizu,H. and Urushihara,H.
            Sexual cDNA in CSM
            Unpublished (1997)
            PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.
            Location/Qualifiers
            source
            1. .554
            /organism="Dictyostellium discoideum"
            /strain="KAX3"
            /db_xref="taxon:44689"
            /cell_type="Gamete"
            /clone_lib="FC"
            /dev_stage="Sexually competent"

BASE COUNT  117 a 36 c 123 g 278 t
ORIGIN

Query Match 4.1%; Score 40.6; DB 19; Length 554;
Best Local Similarity 62.4%; Pred. No. 5.4;
Matches 64; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 203 atagaacagttctaacagtaagtactctcaaaacaaagcatcatggaagacagatgg 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 ACACACAGCTATCCCAACAAAGACGACACACATATATTCACAAACACACACACA 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 263 agtacttgaatacaacaacatggcgacaataacaacaacaataa 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 ACAACAACAACAACAACAACAACAACAACAACAACAACAACA 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
LOCUS    C91401 274 bp mRNA EST 15-APR-1998
DEFINITION Dictyostellium discoideum slug cDNA, clone SSK170, mRNA sequence.
ACCESSION  C91401
NID        93060767
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE     Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA,
            clone lib:SS.
ORGANISM   Dictyostellium discoideum
            Eukaryota; Dictyostellida; Dictyostellium.
            1 (bases 1 to 274)
            Urushihara,H.
            Developmental cDNA in Dictyostellium discoideum
            Published Only in database (1998) In press
            2 (bases 1 to 274)
            Urushihara,H.
            Direct Submission
            Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko

```





```

Query Match          4.0%; Score 39.8; DB 23; Length 560;
Best Local Similarity 57.7%; Pred. No. 8.2;
Matches 71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 875 atgcatttaaatggtcttctgtctatatttattttacatggtctatcaggt 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 ATGAAGTTTCCAATAATTTTTTCTTTTTCCTTTCTTTTCTGTTGTTTAGAT 66

QY 935 ttgcatttaagatttagtacctgtcagatttaaaagaaacaaagttaaaataaaaa 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 AAAAAATTTTGAGAAAAAATAATGAAAAATTTTAAATAAGAGAAAAATGAAAAAATA 6

QY 995 aaa 997
    |||
Db 5 AAA 3

RESULT 14
C91220          574 bp      mRNA      EST      15-APR-1998
Dictyostelium discoideum slug cDNA, clone SSJ7834, mRNA sequence.
C91220
NID 93060586
KEYWORDS EST; EST(expressed sequence tag).
SOURCE Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
clone_lib:SS.
ORGANISM Dictyostelium discoideum
Eukaryota; Dictyostelid; Dictyostelium.
REFERENCE
AUTHORS Urushihara,H.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Published Only in DataBase (1998) In press
REFERENCE 2 (bases 1 to 574)
AUTHORS Urushihara,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
Urushihara, University of Tsukuba, Institute of Biological
Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
(E-mail:qxurushie@bank.dna.afrc.go.jp, Tel.:81-0298-53-4664,
Fax:+81-0298-53-6614)
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
source
1. .574
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SS"
/dev_stage="slug"
BASE COUNT 258 a 76 c 63 g 176 t 1 others
ORIGIN

Query Match          4.0%; Score 39.8; DB 28; Length 574;
Best Local Similarity 60.7%; Pred. No. 8.2;
Matches 65; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 227 acttcaaaaaaagcatcatggaagacagatggagtacttgataacaacaatggcg 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 386 AATTGAAACCAATATAATAAGAACCAACAAACAAACAAACAAACAAACAAACAA 445

QY 287 acataacaacaataatgttacagctcattacgagataattat 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAT 492

RESULT 15
C22974          546 bp      mRNA      EST      27-MAR-1997
Dictyostelium discoideum gamete cDNA, clone FC-AL20, mRNA sequence.
C22974
NID g1922123
KEYWORDS EST; EST(expressed sequence tag).
SOURCE Dictyostelium discoideum (strain:KAX3) Sexually mature Gamete cDNA

```

```

to mRNA, clone_lib:FC.
Dictyostelium discoideum
Eukaryota; mitochondrial eukaryotes; Dictyostelid;
Dictyostelium.
REFERENCE 1 (bases 1 to 546)
AUTHORS Urushihara,H.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
Urushihara, University of Tsukuba, Institute of Biological
Sciences; Ten-noudai 1-1-1, Tsukuba-shi, Ibaraki-ken 305, Japan
(E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
Fax:0298-53-6614)
REFERENCE 2 (sites)
AUTHORS Shimizu,H. and Urushihara,H.
TITLE Sexual-cDNA in CSM
JOURNAL Unpublished (1997)
COMMENT PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
source
1. .546
/organism="Dictyostelium discoideum"
/strain="KAX3"
/db_xref="taxon:44689"
/cell_type="Gamete"
/clone_lib="FC"
/dev_stage="Sexually mature"
BASE COUNT 317 a 110 c 33 g 85 t 1 others
ORIGIN

Query Match          4.0%; Score 39.4; DB 17; Length 546;
Best Local Similarity 57.9%; Pred. No. 10;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 185 agaaaatggcaggcatggtatagaacacagtttcaacagtaagtacttcaaaaagca 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 AAAAAATTGAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 219

QY 245 tcattggcaagacagatggagtacttgataacaacaatggcagacaataacaacaata 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACA 279

QY 305 a 305
Db 280 A 280

Search completed: July 15, 1999, 12:07:21
Job time: 3283 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 15, 1999, 13:56:00 ; Search time 822.9 Seconds  
(without alignments)  
2886.748 Million cell updates/sec

Title: US-09-156-580-1\_COPY\_190\_807  
Perfect score: 618  
Sequence: 1 atggcaggcatggatagaaa.....aacttcgactgggacccact 618

Scoring table: IDENTITY\_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_bal.\*
- 3: gb\_in.\*
- 4: gb\_in.\*
- 5: gb\_ov.\*
- 6: gb\_ov.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pl.\*
- 10: gb\_pr.\*
- 11: gb\_pr.\*
- 12: gb\_pr.\*
- 13: gb\_ro.\*
- 14: gb\_sy.\*
- 15: gb\_sy.\*
- 16: gb\_un.\*
- 17: gb\_un.\*
- 18: gb\_hgt.\*
- 19: em\_ba.\*
- 20: em\_fun.\*
- 21: em\_hum1.\*
- 22: em\_hum2.\*
- 23: em\_in.\*
- 24: em\_in.\*
- 25: em\_or.\*
- 26: em\_ov.\*
- 27: em\_pat.\*
- 28: em\_ph.\*
- 29: em\_pl.\*
- 30: em\_ro.\*
- 31: em\_sy.\*
- 32: em\_un.\*
- 33: em\_vl.\*
- 34: em\_hgt.\*
- 35: em\_sts.\*
- 36: gb\_bal.\*
- 37: gb\_bal.\*
- 38: gb\_pl.\*
- 39: gb\_pl.\*
- 40: gb\_pr.\*
- 41: gb\_pr.\*
- 42: gb\_pr.\*
- 43: gb\_sts.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query  
No. Score Match Length DB ID Description

1	89	14.4	1071	9	ATU38946	U38946 Arabidopsis
2	89	14.4	1071	39	ATU38946	U38946 Arabidopsis
c 3	76.2	12.3	99687	18	AC005956	AC005956 *** SEQUE
c 4	70.4	11.4	93845	9	ATAC004684	AC004684 Arabidops
c 5	70.4	11.4	93845	39	ATAC004684	AC004684 Arabidops
6	49	7.9	5931	3	DDU32174	U32174 Dictyostell
7	46.8	7.6	1200	8	MISGVA	V00705 Yeast mitoc
8	46.8	7.6	1200	8	YSCMTVAR1	J01525 Yeast (S.ce
9	46.8	7.6	1200	38	MISGVA	V00705 Yeast mitoc
10	46.8	7.6	1200	38	YSCMTVAR1	J01525 Yeast (S.ce
11	46.2	7.5	1083	8	PETZFDB2	D26084 Petunia mRN
12	46.2	7.5	1083	38	PETZFDB2	D26084 Petunia mRN
13	44.6	7.2	3041	8	PETZFDB1	D26083 Petunia zin
14	44.6	7.2	3041	38	PETZFDB1	D26083 Petunia zin
15	44.2	7.2	12668	3	AF079445	AF079445 Dictyoste
16	44.2	7.2	1453	3	DDI039PPR1	AF018638 Dictyoste
17	43.8	7.1	7467	3	PFSC04099	AL010282 Plasmodiu
18	43.8	7.1	5100	8	YSCPH081	D13228 Yeast PHO81
19	43.8	7.1	5100	38	YSCPH081	D13228 Yeast PHO81
20	42.2	6.8	3478	8	DDIADCYG	M87278 Dictyostell
21	42.2	6.8	3478	38	DDIADCYG	M87278 Dictyostell
22	42	6.8	4374	3	AF045453	AF045453 Dictyoste
23	41.8	6.8	235	3	PFPPF297	X53020 P.falciparu
24	41.4	6.7	165	43	G37866	G37866 emaa2 Plasm
25	41.2	6.7	6265	3	PFSC03080	AL010153 Plasmodiu
26	41.2	6.7	2107	8	DDIADCYA03	L05498 Dictyostell
27	41.2	6.7	2107	38	DDIADCYA03	L05498 Dictyostell
28	41.2	6.7	221	43	G37927	G37927 c3m61 Plasm
29	41	6.6	93419	18	AC002415	AC002415 *** SEQUE
c 30	41	6.6	231	43	PCU64653	U64653 Paracharter
31	40.8	6.6	7133	3	AF024654	AF024654 Dictyoste
32	40.8	6.6	3887	3	AF029726	AF029726 Dictyoste
33	40.8	6.6	2239	3	DDU07817	U07817 Dictyostell
34	40.8	6.6	6115	3	DDU20432	U20432 Dictyostell
35	40.8	6.6	1191	9	SCE010480	AJ010480 Saccharom
36	40.8	6.6	1191	39	SCE010480	AJ010480 Saccharom
37	40.6	6.6	5108	2	BBU44914	U44914 Borrelia bu
38	40.6	6.6	1690	6	I52144	I52144 Sequence 24
39	40.6	6.6	1700	8	S79359	S79359 arabinogala
40	40.6	6.6	5108	37	BBU44914	U44914 Borrelia bu
41	40.6	6.6	1700	38	S79359	S79359 arabinogala
42	40.4	6.5	60232	2	AE001272	AE001272 Lactococc
c 43	40.4	6.5	41076	3	CER02D5	Z78015 Caenorhabdi
44	40.4	6.5	1640	8	AB000453	AB000453 Petunia h
c 45	40.4	6.5	232854	18	CEY40H4	AL022573 Caenorhab

ALIGNMENTS

RESULT	1	ATU38946	1071 bp	DNA	PLN	29-NOV-1995
LOCUS		Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.				
DEFINITION		Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.				
ACCESSION		U38946				
NID		91079668				
KEYWORDS		thale cress.				
SOURCE		Arabidopsis thaliana				
ORGANISM		Eukaryotes; mitochondrial eukaryotes; Viridiplantae;				
REFERENCE		Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;				
AUTHORS		Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.				
TITLE		1 (bases 1 to 1071)				
JOURNAL		Sakai, H., Medrano, L.J. and Meyerowitz, E.M.				
MEDLINE		Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries				
REFERENCE		Nature 378 (5553), 199-203 (1995)				
AUTHORS		96069740				
TITLE		2 (bases 1 to 1071)				
JOURNAL		Sakai, H., Medrano, L.J. and Meyerowitz, E.M.				
FEATURES		Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech, Pasadena, CA 91125, USA				
		Location/Qualifiers				

```

source      1. .1071
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/chromosome="3"
join(1. .68,175. .202)
/genes="sup"
1. .1071
/genes="sup"
join(1. .68,175. .1071)
mRNA
/genes="sup"
69. .174
intron
/genes="sup"
203. .817
CDS
/genes="sup"
/standard_name="SUPERMAN"
/notes="Allele: Ler wild type; C2H2 zinc finger protein"
/codon_start=1
/function="flower-specific gene controlling the boundary
of the stamen and carpel whorls"
/product="SUPERMAN"
/db_xref="PID:g1079669"
/translation="MERSNSIELRNSFYGRATSPWSYGDYDNCODHDYLLGFSWPP
RSYTCFCKRERSAALGCHMNVHRRARLRLOQSPSSSTPSPYPNPNYSYSTM
ANSPPHSPKLLFPILSPSPRYRAGLRISLSPKSKHTPENACKTKKSSLVGE
ATRFTSKDACKILRNDEIISLEIGLINESQDLDLELRGFA"
818. .1071
/genes="sup"

3' UTR
BASE COUNT 326 a 226 c 183 g 336 t
ORIGIN

Query Match      14.4%; Score 89; DB 9; Length 1071;
Best Local Similarity 68.9%; Pred. No. 9.7e-10;
Matches 122; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 124 agctcattacagagataatttgaaatgaagatcatttacttgggtgactattctcttgg 183
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 AGCTATGGAGATTATGATAATTGCCAACAGAGATCATGATTATCTTCTAGGGTTTTCATGG 328
QY 184 cctccaagatcttatcatagcttttataaagggaatttagactctcctcaagctctt 243
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 329 CCACCAAGATCCTACATTCGACGCTTCGCAAAAGGGAATTCAGATCGGCTCAAGCACTT 388
QY 244 ggtggacacatgaattgttcataagaagatagagccatttttgagacaatcaccacct 300
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 389 GGTGGCCACATGAATGTTCCACAGAAGACAGACAGCAAGACTCAGATTACACAGTCT 445

RESULT 2
AC005956/c
LOCUS
DEFINITION
ATU38946 1071 bp DNA PLN 29-NOV-1995
Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.
ACCESSION
U38946
NID
g1079668
KEYWORDS
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 1071)
AUTHORS
Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
TITLE
Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries
JOURNAL
Nature 378 (6553), 199-203 (1995)
MEDLINE
96069740
REFERENCE
2 (bases 1 to 1071)
AUTHORS
Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
TITLE
Direct Submission
JOURNAL
Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech,
Pasadena, CA 91125, USA
FEATURES
Location/Qualifiers
1. .1071
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/chromosome="3"
join(1. .68,175. .202)
/genes="sup"
1. .1071
/genes="sup"
join(1. .68,175. .1071)
mRNA
/genes="sup"
69. .174
intron
/genes="sup"
203. .817
CDS
/genes="sup"
/standard_name="SUPERMAN"
/notes="Allele: Ler wild type; C2H2 zinc finger protein"
/codon_start=1
/function="flower-specific gene controlling the boundary
of the stamen and carpel whorls"
/product="SUPERMAN"
/db_xref="PID:g1079669"
/translation="MERSNSIELRNSFYGRATSPWSYGDYDNCODHDYLLGFSWPP
RSYTCFCKRERSAALGCHMNVHRRARLRLOQSPSSSTPSPYPNPNYSYSTM
ANSPPHSPKLLFPILSPSPRYRAGLRISLSPKSKHTPENACKTKKSSLVGE
ATRFTSKDACKILRNDEIISLEIGLINESQDLDLELRGFA"
818. .1071
/genes="sup"

3' UTR
BASE COUNT 326 a 226 c 183 g 336 t
ORIGIN

Query Match      14.4%; Score 89; DB 39; Length 1071;
Best Local Similarity 68.9%; Pred. No. 9.7e-10;
Matches 122; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 124 agctcattacagagataatttgaaatgaagatcatttacttgggtgactattcttcttgg 183
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 AGCTATGGAGATTATGATAATTGCCAACAGAGATCATGATTATCTTCTAGGGTTTTCATGG 328
QY 184 cctccaagatcttatcatagcttttataaagggaatttagactctcctcaagctctt 243
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 329 CCACCAAGATCCTACATTCGACGCTTCGCAAAAGGGAATTCAGATCGGCTCAAGCACTT 388
QY 244 ggtggacacatgaattgttcataagaagatagagccatttttgagacaatcaccacct 300
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 389 GGTGGCCACATGAATGTTCCACAGAAGACAGACAGCAAGACTCAGATTACACAGTCT 445

RESULT 3
AC005956/c
LOCUS
DEFINITION
AC005956 99687 bp DNA HTG 04-DEC-1998
*** SEQUENCING IN PROGRESS *** Arabidopsis thaliana 'MITSUI' P1
'MHK10', genomic sequence near marker 'EG05C12'; HTGS phase 2, 1
ordered pieces.
ACCESSION
AC005956
NID
g3962504
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (bases 1 to 99687)
AUTHORS
Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M., Shen,M.,
Ronning,C.M., Rounsley,S.D., Fraser,C.M., Somerville,C.R. and
Venter,J.C.
TITLE
Arabidopsis thaliana 'MITSUI' BAC 'MHK10' genomic sequence near
marker 'EG05C12'
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 99687)
AUTHORS
Rounsley,S.D. and Lin,X.
TITLE
Direct Submission

```



repeat_region	QSLTPCSVENHYSEKTYAKDTLDEPKRIAKLTSVSGTGCHKKKXVSEGSNRRKRNDLHRLFPNGLPDGTETLAYVYKTOISPSQFEAHAGMAARQPYRHIIFISGLSHUDIAMSLANGHVITTDSDMCCSGICGGDGLLLCAGCPQAFHTACLKAFQSMPEGTWFCSSCNDGPISKKKATITDPSGNARPIVIRLSVRKAPESDIGCVFCRSHDFSIGKDDRTVILCDQCEKREYHVGCLRENGCFDLEIPOEKWPCSCSRIHTAVONSVSCGPOTLPPLLDMICRDREKGIETDGDVVEVRIILSGKSRYPPEHLLPRAAVIFRECDDPIVAKSGRLPVMVYIGNLSQGGMYCLVLIVNSLVSAALLRIFGQVEALPIVATSVREYQGRYFOGLYACVENLLSNVENLVLPAAEEAESIMTKKFGTKMSDQQLQEQYQEVQLTIFPGTSMLEKKVPKATITGLSESTILI"	
	6010..6135	
misc_feature	/rpt_family="(GGAA)n"	
	6023..6202	
repeat_region	/note="exon predicted by xgrail, quality marginal_shadowexon"	
	6179..6236	
mRNA	/rpt_family="POLY_A"	
	join(<6647..6702,7161..7304,7322..>7499)	
gene	/gene="F13M22.3"	
	<6647..>7499	
CDS	/note="predicted by genscan"	
	join(6647..6702,7161..7304,7322..7499)	
CDS	/note="hypothetical protein"	
	/codon_start=1	
mRNA	/db_xref="PID:g3236236"	
	SLAPEAFSAIFPSSESVFSRRHIGSAGSLGLVGTGESPEVEDGPRCAAGFCNCCGSRRRKDGPLGGRBPLI"	
gene	complement(join(<7761..7931,8019..8065,8154..8238,8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))	
	/gene="F13M22.4"	
CDS	complement(<7761..>9886)	
	/gene="F13M22.4"	
CDS	complement(join(7761..7931,8019..8065,8154..8238,8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))	
	/gene="F13M22.4"	
repeat_region	/product="putative ribitol dehydrogenase"	
	/codon_start=1	
repeat_region	/db_xref="PID:g3236237"	
	IGLEARVLARMGAHVIIIAARNPKANESKEMIQMNPARNVDYIQLDVSXKVSFVDFALNPLNLIINAGMFCPPFLKLEDGESQATNHHGLNLNLLDKMSKVSFRESQOGRVINSASHIATYISGIRKFGINDPAGYSERRAYGQSKLNLHNSALSRRLQOEGVNITINSVHPLVTNLFYSGFSMKVFRAMTELFWKNIPOGAATTCYVALHPLEGVGTGFGDCNIVAPSKPATNNSLADKLWDFSVFLDTSIK"	
repeat_region	8658..8684	
	/rpt_family="AT-rich"	
repeat_region	complement(9842..9915)	
	/rpt_family="(GA)n"	
misc_feature	10503..10670	
	/note="exon predicted by xgrail, quality marginal_shadowexon"	
repeat_region	10725..10746	
	/rpt_family="AT-rich"	
mRNA	complement(join(11416..11526,11608..12116,12551..12628,12737..13604))	
	/gene="F13M22.5"	
gene	complement(11416..13604)	
	/gene="F13M22.5"	
CDS	complement(join(11610..12116,12551..12628,12737..13522))	
	/gene="F13M22.5"	
repeat_region	/codon_start=1	
	/product="putative ARF1 GTPase activating protein"	
repeat_region	/db_xref="PID:g3236238"	
	KHRLGVHSISFVSRYTSDSWSEIQIKMDAGGERLNUNFLAQIGISKETDIISKYNSNAASVYRDRQAALAEQRWRDPPIVRESVGGGLMKRPPLSQGGGRDSGNGWMDNDLDFSRDTRMNRNAGDVGSGRGAPAKSKSSEDIYSRQLEASAANKESFTAKRMAENESKPEGLPPSQGKVYGGSPGAPRNSNQSGGDFVSMSEGGFRLSLVAASAAANVQOTGMEETSKNKEGGLDQVSETVYVYVASKTEIGQRTWIGIMGYMAIASQKVEEFTKEASTNQKNTEGNIQNGIKNTANSFGGSSQSSGHNNNSWDDWGEBENNSKKEAAPKYSTNSDDDDGGWAGWDDNDKADDDFYYPQASDKKSVGHNGKSDTAW	
repeat_region	TGGGFL"	
	11697..11757	
repeat_region	/rpt_family="(CAT)n"	
	13716..13952	
misc_feature	/note="exon predicted by xgrail, quality excellent_shadowexon"	
	complement(14172..14309)	
misc_feature	/note="exon predicted by xgrail, quality marginal"	
	join(14707..14911,15566..15765,15858..15964,16057..16189,16311..16602,16755..17294)	
mRNA	/gene="F13M22.6"	
	14707..17294	
gene	/note="identical to GB:U40269"	
	/gene="F13M22.6"	
CDS	join(15634..15765,15858..15964,16057..16189,16311..16602,16755..17182)	
	/gene="F13M22.6"	
CDS	/codon_start=1	
	/product="origin recognition complex protein"	
mRNA	/db_xref="PID:g3236239"	
	/translation="MEDIENTEDEVGFSRNYFLAKELGGASKRAHKLSDIHIVDEQELRETASTRIEMKHSEISLMSDYKTMYSKWVEELRCGGLLMYGSKKALVEDFASASLTDYSVVVINGYLPDSVNLKQVLLALAELLSELKCKRKSXGQETPFPSRMDILSLFHGPQSGDKDFICVVVHNIDGPAIRDPEQQTLARLSSSHRLVASIDHVNAPLLWDKKVHKQFNWLMHHVPTFAPYINVEGVFFPLVLAQGSTAQTAATAIVLQSLTPNGQNVFKILABYQLSHPEDGMPDLDLISASRERFFVSSQVTLNSHLTEFKDHELKVTNRNSDGOECLNIPITSDAIROLLDLNQ"	
mRNA	complement(<17532..>19016)	
	/gene="F13M22.7"	
gene	complement(<17532..>19016)	
	/gene="F13M22.7"	
CDS	complement(17532..19016)	
	/gene="F13M22.7"	
CDS	/note="unknown protein"	
	/codon_start=1	
mRNA	/db_xref="PID:g3236240"	
	/translation="MENHPSTLLSMDSSASSHEELDELMNNRSLSGPPDINLPLSAERSPPPPWNLDACDILDVGLGSAYETENYMSVPKGRKCAKRVDSINGANWFFSFYKPALNEKSKAKIVRDSNGISGFDKSLDLVLYQHDNMNMYVFKERPENALCKMQLRSYMNCHSGROGDLFPFSVEGKVSRRMQRKHRYGLSNQCCHGIELPLPLNLTCLDEERKRWMLTGRDLNFTIPEASDFGSWRNLNPTDFELSRPTPSLKPNANH	
misc_feature	SKKLLSGNLSTQPSNHSNGEATDLSPPSHKKRDLFNGIHIEECCLTVPQPPVIBAHQNEPLTWSNEFTGAMKNYGPVTAATKIYDEDEGYLIISLPFVDLNSVKVSWRNLTHGIIKVSCLSTSRVPFKRRHRTFKLTDASEHCPGPFVREIPLSNRIPEDANIBAYYDGGSVLEILVPLKRAPGPEHVRVCLRPNLGGNDMLT"	
	complement(19045..19091)	
misc_feature	/note="exon predicted by xgrail, quality marginal"	
	19455..19545	
misc_feature	/note="exon predicted by xgrail, quality excellent_shadowexon"	
	join(<20811..21507,21667..21805,22028..22434,22931..23006,23118..23233,23331..>23788)	
mRNA	/gene="F13M22.8"	
	<20811..>23788	
gene	/gene="F13M22.8"	
	join(20811..21507,21667..21805,22028..22434,22931..23006,23118..23233,23331..23788)	
CDS	/gene="F13M22.8"	
	/codon_start=1	
CDS	/product="putative zinc finger protein"	
	/db_xref="PID:g3236261"	
CDS	/translation="MNFNTT"SFSGSGGVVVVPTATTVPPTDFGTTITSNSTFIIGPPPEPAPPSIDITPLKLIFVIAFVAPALVYALFENGPCSSRRNSSSRTSSSSDDTTHAVTDPIITETTVTSESQGFHKDTHSKEIGNECSVCLAVTDSDELRLSOLSKHAFVLCIETWLADHNCPICTRDVSVKQOTEAPNPVNVNNGVNSGGNRVSAT	
	SRDDWRQGLPDASTTVSVSDVSDVSDQAHPTVYEDLLILGRLLIHTRLPRTL	
CDS	SFCSSRSDSSGLAPLNATKSTIPRFAYLTGTGDKGRVKRLLAIIHPRNYIL	
	LKLDLEASDERMELAKYVRKDDKFNVMGLADLVTEKGTPLASTLHGVAIILK	
CDS	RAKDDWFINLSADYPLMPQDDILHIESYLPYLNFTHTSNIGWKNORARPILL	
	PCFYHLKSGYVFWAKERSLPASFKLFWGSSVALTRPFLFCINQWNLPMVGY	
CDS	TNFLSSGEGYFQTVCCNKDQNTVNHDLHYTKWDPLOQRTLVNTVENFRDMVSG	
	PFARSGEDDLVLDKIDIELLGQTDTGLELTKTPDVVVKPTVSWKRLEKLMVRLLDHENF	
CDS	RAKQK"	

```

repeat_region complement(20872..20971)
repeat_region /rpt_family="(TGG)n"
repeat_region complement(21518..21543)
repeat_region /rpt_family="AT-rich"
misc_feature complement(24200..24291)
repeat_region /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region complement(25078..25159)
repeat_region /rpt_family="(TA)n"
misc_feature join(<25358..25408,25622..>26563)
misc_feature /gene="F13M22.9"
misc_feature <25358..>26563
misc_feature /gene="F13M22.9"
misc_feature join(25358..25408,25622..26563)
misc_feature /gene="F13M22.9"
misc_feature /codon_start=1
misc_feature /product="putative zinc finger protein"
misc_feature /db_xref="PID:g3236241"
misc_feature /translation="MVFSIIQAYLDSSNNQQAPSPNYNHGDTGSANGGHVLRPOLQ
OPOQOQPHPGSGGGGGGIRAGSMVDRARQANVALPEALKCPCESTNTKFC
YFNYSLTQPRHECTCRXYWTRGALRNVPVGGCRNRRTKSNNSNNSTATSN
TSFSGNSTSTILSSHYGNQESILSOILSPARLMPNTYHGLDLSNTKIDNMS
LLNYGLSODRSHIMGASGSLSCVDWRSASYHQOSSMGGLDSDSNPNPSANG
FYFSPRITSASISALASQSSVKVEDNPYKWNVANGCSSWNDLSAFGSSR"
repeat_region 25691..25728
repeat_region /rpt_family="(CAG)n"
repeat_region 26002..26054
repeat_region /rpt_family="(CAA)n"
misc_feature complement(27038..27122)
repeat_region /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region 28179..28208
repeat_region /rpt_family="AT-rich"
misc_feature complement(28259..28328)
repeat_region /note="exon predicted by xgrail, quality
marginal_shadowexon"
repeat_region complement(28897..28980)
repeat_region /rpt_family="(TA)n"
misc_feature 29086..29133
repeat_region /rpt_family="AT-rich"
misc_feature complement(29945..30023)
repeat_region /note="exon predicted by xgrail, quality good"
misc_feature complement(join(<30476..30515,30659..30858,31168..>31269))
misc_feature /gene="F13M22.10"
misc_feature complement(<30476..>31269)
misc_feature /gene="F13M22.10"
misc_feature complement(join(30476..30515,30659..30858,31168..31269))
misc_feature /gene="F13M22.10"
misc_feature /codon_start=1
misc_feature /product="putative ribosomal protein L36"
misc_feature /db_xref="PID:g3236242"
misc_feature /translation="MTTAVKTYGLVGLKGVHVTRELPNPNRKRKTSKRTIFIR
KLIREVAGMAPYKRYKITELLKVGDKRALKVAKRKLGTGTHKRAKRREEMSSVLRKMS
LGGAABAEKMK"
misc_feature complement(31377..31401)
misc_feature /note="exon predicted by xgrail, quality marginal"
misc_feature complement(32006..32036)
repeat_region /note="exon predicted by xgrail, quality marginal"
repeat_region 32044..32161
repeat_region /rpt_family="(GAAA)n"
repeat_region complement(32282..32354)
repeat_region /rpt_family="(GAAAA)n"
repeat_region 32653..32718
repeat_region /rpt_family="POLY_A"
misc_feature complement(join(<32935..33310,33630..>33739))
misc_feature /gene="F13M22.11"
misc_feature complement(<32935..>33739)
misc_feature /note="predicted by genscan"
misc_feature /gene="F13M22.11"
misc_feature complement(join(32935..33310,33630..33739))
misc_feature /gene="F13M22.11"
misc_feature /note="hypothetical protein"
misc_feature /codon_start=1
misc_feature /db_xref="PID:g3236243"
misc_feature /translation="MEIGNVADNVLNLREAEHRHRYFDYFIDEILWLSFLGNSET

```

```

SLKEEVTLGCVKETREECSDDKLVLDICLKRRLDEEEVEEVEEENDEGFKTPTRP
ENRPIPIVRECPAPMKRSESEMERGRTMYCRRRLSELPEDDDVNSITDLQWTTMTIK
K"
repeat_region complement(33140..33196)
repeat_region /rpt_family="(CAA)n"
repeat_region 33357..33398
repeat_region /rpt_family="(GAA)n"
repeat_region 33469..33489
repeat_region /rpt_family="AT-rich"
misc_feature complement(34248..34296)
misc_feature /note="exon predicted by xgrail, quality
marginal_shadowexon"
misc_feature 35206..35258
misc_feature /note="exon predicted by xgrail, quality good"
repeat_region complement(35273..35341)
repeat_region /rpt_family="(GGA)n"
misc_feature join(35801..35886,36023..36416,36517..37130,37242..37322)
misc_feature /gene="F13M22.12"
misc_feature 35801..37322
misc_feature /note="identical to GB:U29480"
misc_feature /gene="F13M22.12"
misc_feature join(35827..35886,36023..36416,36517..37130,37242..37307)
misc_feature /gene="F13M22.12"
misc_feature /codon_start=1
misc_feature /product="actin 3 protein"
misc_feature /db_xref="PID:g3236244"
misc_feature /translation="MADGDIQPLVCDNGTGMVKAGFAGDDAPRAVFPFIVGRPRHTG
VMVGOKDAYVGEAQSRIKILTYPIEHGIVNMDDMEKIWHHTYINELRVAPEE
HPILTEAPLNPKANRKMTOIMETFNAPAYVIAQVLSVSGRTTGIVLDSGDG
VSHVPIYEGVALPHAILRLDLADRLTDALMKILTERGYSTTAEREIVRDLKEKL
CYIALDYEOELETAKTSSSVKNEVELPDGQVITIGSERFRCPEVLYQPSMIGMENAGI
HETYSIMKCDVDILKLDLYNLVSGGTTMFPFGIADRSKEITLALAPSSWKIKVAP
PERKYSYMWIGGSVILASLSTFOQMIAKAEYDESGPSIVHRKCF"
repeat_region 37700..37817
repeat_region /rpt_family="(CAAAA)n"
repeat_region complement(<37864..>38967)
repeat_region /gene="F13M22.13"
repeat_region complement(<37864..>38967)
repeat_region /gene="F13M22.13"
repeat_region complement(37864..38967)
repeat_region /gene="F13M22.13"
repeat_region /codon_start=1
repeat_region /product="myb-related protein"
repeat_region /db_xref="PID:g3236245"
repeat_region /translation="MKERQWSEEDALLRAYVRQFGPREWHLYSERMKNPLNRDAKS
CLERWNYLPGIKKNGKSLTEEQRLVIRLOEKHGKWKKIAAEVPGRTAKRLGKWEV
FKEQOREEKESNKRVEPIDESKYDRILEFAEKLKRSNVVPAATAAATVVMANS
GGFLHSEQQVQPPNPVPPWLATSNNGNVVAPPSVTLTSPSTVAAAAPQPPIPWL
QQQOPERAENGPGGLVLCGSMPPSCGSESVFLSELVECCRELEEGHRAADHKEAA
WRLARLEQLSEKTCQREKMEIEAKMKALREEQNAMEKIEGEYREQLVGLRRDA
EAKQKLDQWTSRHIRLTFLFQQMGCLDRP"
repeat_region complement(39531..39561)
repeat_region /rpt_family="AT-rich"
repeat_region complement(40737..40818)
repeat_region /rpt_family="(TAA)n"
repeat_region complement(41089..41112)
repeat_region /rpt_family="POLY_A"
repeat_region complement(41190..41213)
repeat_region /rpt_family="POLY_A"
repeat_region complement(41283..41310)
repeat_region /note="exon predicted by xgrail, quality
marginal_shadowexon"
repeat_region complement(41558..41602)
repeat_region /rpt_family="AT-rich"
repeat_region complement(41966..42068)
repeat_region /rpt_family="(GAAAA)n"
repeat_region complement(42281..42325)
repeat_region /rpt_family="AT-rich"
repeat_region 42889..42968
repeat_region /note="exon predicted by xgrail, quality excellent"
repeat_region 43132..43154
repeat_region /note="exon predicted by xgrail, quality good"
repeat_region 43371..43479

```

mrna /rpt\_family="TAAAA)n"  
complement(join(43750. .44452,45240. .45558,45719. .45963))  
/gene="F13M22.14"  
gene complement(43750. .45963)  
/gene="F13M22.14"  
CDS complement(join(44143. .44452,45240. .45558,45719. .45878))  
/gene="F13M22.14"  
/codon\_start=1  
/product="putative expansin protein"  
/db\_xref="PID:g3236246"  
/translation="MTATFRVGLWLAVTASFLLTATNAKIPGVYSGGPWQNAHATFY  
GGSDATMGACGYGLYSGQYGVNTAAALSTALENNGFCGACFEIKCTDDRWCV  
GNPSILVTAINFCPNPFAQSDGGWCNPPREHFDLAMPFLKIGLIRAGLIVPVSYR  
VPCRIIGIRTVTFNGFIYFNLVITNAGADINGVSKGSDTDMWRMSRWGQWQS  
NAVIGOSLFRVTRDRSRSTSMNVAPATWQGTSGKNFRV"  
44795. .44903  
repeat\_region /rpt\_family="AT\_rich"  
repeat\_region 46287. .46312  
misc\_feature /rpt\_family="AT\_rich"  
complement(47301. .47444)  
/note="exon predicted by xgrail, quality  
excellent\_shadowexon"  
48559. .48589  
repeat\_region /rpt\_family="AT\_rich"  
mrna <48689. .>50845  
gene /gene="F13M22.15"  
CDS /gene="F13M22.15"  
48689. .50845  
/gene="F13M22.15"  
/codon\_start=1  
/product="SCARECROW-like protein"  
/db\_xref="PID:g3236247"  
/translation="MITEPSLTIGSGMVNRNLGLDPQSSHSFTPTVLYDGFVNL  
SSDHINVAAPENSVFIREEEEDPADDFSDAVLGVISOMLNEEDMDKVCMLQ  
ESLDEAERSLIEAIKCKYPPSPERNLAFARNSENLDRAVPGNYTGGCIGFNGG  
IKPUSSTDFDRNPQSCSILSVPSQNLITTYDGDIDESSKNRNENHOSVLFRE  
IEEANRPENELIVFRENCVSKARKNSRDEICVEERSKSLPAVFGEDILRSD  
VVDKLVHVPGGESGEMFNALVDLVKGVKKKASDAQGGKRARGRGGRGGGG  
QNGKKEVVDLSLIIHCAQAAVADDRACAGLLQKIRLHSTPFGDGNORLAHCFANGL  
EARLAGTSOIIYKIVSKPSRAAVLKAHOLFACCPFKLSYFTNKTIRDLVGNQ  
RVHVDGILYGFOWPPLIHRFSWYSGPKVITGIEFPQGFPAQKVETGRLAAY  
AKLVGFPEFYKAIAKKWDAIQLEDLDIDREITIVNCLYRAENLHDESIVSCRDIV  
LNLGKINPOLFYINGYAYNAPFFVTRFEALFHSIFDMLETIVPREDEERMF  
EMEVFGREALNVIACEGWVERPETYKQHVRAMRSLGVQVPDPSIMKTSLHKVHT  
FYHKDFVIDODNRWLQGWKGRVTMALSVNKPESKA"  
49591. .49709  
repeat\_region /rpt\_family="GAA)n"  
repeat\_region 49635. .49689  
misc\_feature /rpt\_family="GGA)n"  
misc\_feature 51112. .51155  
/note="exon predicted by xgrail, quality good"  
complement(51301. .51387)  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
mrna complement(join(51428. .51666,51782. .51905,51988. .52049,  
52131. .52264,52306. .52342,52427. .52698,52779. .>53043))  
/gene="F13M22.16"  
gene complement(51428. .>53043)  
CDS /gene="F13M22.16"  
52131. .52264,52306. .52342,52427. .52698,52779. .53043))  
/gene="F13M22.16"  
/note="unknown protein"  
/codon\_start=1  
/db\_xref="PID:g3236248"  
/translation="MAWMTTITTTFFHPLLPANTYKSGAVASSVSVPRSSSIQFRSL  
VSDSTICGSPKSGKNRRSVTVSAAATTEPLTVLTGAGRTGQIVYKKLERSQ  
FVARGLVTKESKEKINGEDEVFIGDIRDTASIAPEVIGDALVILTSAVQPKMGPF  
PSKGGRPFEFFDGCAYPEQVDWIGQKNOIDAGVDDGLSFVVTAKAGVQKIVLVGSM  
GCTNINHPISAGNANILVMKRAEOYLADSGIPYTIIRAGGLQDKDGGIRELLVCKD  
DELLETETRTIARADVAEVCVQALQLEEARFKALDILASKPEGTGTPTKDFRALFTQVT  
TKF"

mrna complement(join(<54350. .54129,54196. .54289,54371. .54453,  
54545. .54596,54711. .54868,54958. .>56602))  
/gene="F13M22.17"  
gene complement(<54350. .>56602)  
/note="predicted by grail"  
/gene="F13M22.17"  
CDS complement(join(53450. .54129,54196. .54289,54371. .54453,  
54545. .54596,54711. .54868,54958. .56602))  
/gene="F13M22.17"  
/note="hypothetical protein"  
/codon\_start=1  
/db\_xref="PID:g3236249"  
/translation="MDGNKLRDKMTMNWEGLGDFEDEDDRFETEDHRLSALTTF  
DNANASSDEDEDFDCRLSFSASVSLTASRKRFTPAMSPDPTDWAAPGISERR  
RRLHGMGLAKNDKMSAVSIRRVSNAPVVSNGEEKMKMKNNHVEDQDHVPVM  
LARSSESDIERFLIEKRKEILGSKQRLTRTYSTICPTRICQVOTPTROTPA  
VCGNKALRGGEALTSMVSNARVGAFFLKNLDGKEFVNEYVDGWMNRSLDLOT  
GKOLTEEFKCVGYSVPVKELMRRENVRINPELMDLARKFNSYLSKSVRLSKRGA  
ALLKNIKVAHSSMLRVADKVDSDGTSPPKKGDHKKHAKANEWVKVRPTGSKYKELS  
ALHMCQEIQAHEGAVMTIKFSQDAHYLASGADRVHVMEVQCELSMNEGSLTPIH  
PSLCDSSGNEITVVEKKKGKGGSSRRHNHIDYVHVPEVFSFSDKPVCSLGHDA  
ILDLSWSKSOLLSSSMKDTVRLWDIETCLKLFADNDYVTCIOFSPVDENVFLSGS  
LDAKIRIWSIQDRHVWMSDLHEMVTAACTYPDGALIGSHKGIQRAVDTEDCKLSQ  
TNQIDVQSNKKSQAKRKITSFQSPVNPSEVLVTSADSRILDGSEVHKFKGFRNT  
CSQLSASYSDQGYKIIICASEDSQVYLWNKDFHRTSLTITQSHHEFHCKDVSAAVPMH  
GHVRGPPPVQIHSKRHSKRISTSSQPSSTISSPTKEETSATGPTTSNRNKSGLPMP  
PKKAATKSQIQPEEEAGPELGSSESFRSMNSSEQHSRFGSPSINTSRSSWSWF  
DSGGHGPTQIPTAWGMVLVTATVHGEIRSYQNGFLPRIGROTTLF"  
complement(56145. .56194)  
repeat\_region /rpt\_family="GAA)n"  
misc\_feature 56686. .56744  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
misc\_feature complement(56743. .56808)  
/note="exon predicted by xgrail, quality marginal"  
misc\_feature complement(57092. .57128)  
/note="exon predicted by xgrail, quality marginal"  
misc\_feature 57429. .57497  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
misc\_feature complement(57759. .57834)  
/note="exon predicted by xgrail, quality excellent"  
mrna complement(join(<57887. .58388,59719. .59846,60078. .60132,  
60218. .60291,60384. .60527,61543. .>61581))  
/gene="F13M22.18"  
gene complement(<57887. .>61581)  
/note="predicted by genefinder"  
CDS /gene="F13M22.18"  
60218. .60291,60384. .60527,61543. .61581))  
/gene="F13M22.18"  
/note="hypothetical protein"  
/codon\_start=1  
/db\_xref="PID:g3236250"  
/translation="MPVRVESNTPAQAFSGTQNVSNQKKEAWRVNVOIQGDLDEHG  
YLCGTMEALNVPMDPTVITFWEGEIVDGKNTFTYKWEATREDDMHWSKFPSP  
LQWRKQFYFVNVGTDCGLTIAGFYVCFSCDSGISGFYDPNSSFHHMISKSNVL  
KWEVEVSKRKRFOTQSDLSLLPLSKTKTCFANVACSENTNGNSIDETYSMSYVN  
SNTSMCNKDNIENKESSSCGCKDKMISFESHLDIYITGTONLDFSEKVIENIYLIDE  
QEEEDADAGCCSSNAARFVLSRGRVTWNOGKNRILY"  
58770. .58795  
repeat\_region /rpt\_family="GAA)n"  
repeat\_region 59165. .59234  
/rpt\_family="AT\_rich"  
repeat\_region 59249. .59312  
/rpt\_family="AT\_rich"  
repeat\_region complement(59446. .59478)  
/rpt\_family="AT\_rich"  
misc\_feature complement(61713. .61851)  
/note="exon predicted by xgrail, quality good\_shadowexon"  
mrna join(<62177. .62406,62814. .63035,63131. .63212,63337. .63399,  
63504. .63642,63940. .63995,64072. .64145,64247. .64394,  
64470. .64652,64757. .64844,64948. .65058,65257. .65354,



65591. .65708,65788. .65882,66002. .66108,66213. .>66306)  
/gene="F13M22.19"  
<82177. .>66306  
/gene="F13M22.19"  
join(62177. .62406,62814. .63035,63131. .63242,63337. .63399,  
63504. .63642,63940. .63995,64072. .64145,64247. .64394,  
64470. .64652,64757. .64844,64948. .65058,65257. .65354,  
65591. .65708,65788. .65882,66002. .66108,66213. .66306)  
/gene="F13M22.19"  
/codon\_start=1  
/product="putative phosphoribosylaminoimidazole  
carboxylase"  
/db\_xref="PID:g3236251"  
/translation="MLLKQSSAAVLVVGNTTPVLHTSRSTYRVGFFPVTRTQSQSL  
TWANQKLPFTSSGLKNTASVPCSSHDASPISENREKHNHGVSEKIVGVJGGQGLG  
RMLCQAASLAIKMILDPKNCSSASHSDVSDFSATVEFAKRCVGLTVIEIE  
HYDVTLEKLEQGVDCQPKASTIRIIODKYMOKVHFSQHGLPLPEFMEISIDEGARK  
ACELFGYPLMIKSKRLAYDGRGNANVANNODELSAVTAGLGSRLYTEKWAFFKEL  
AVIVARGDGSVCYPVNETIHRDNICHTIVKAPADVPKINKLATDVAQKAVGSLEGA  
GVFAVELFTEDSQILLNEVAPRPNHSGHOTIECCYTSQFOHLRAVGLPLGDFSMR  
TPASIMNLTGDDVDIGAGEFKLAHRLIALCALCPGASVHWYDKPEMKRKHGHIT  
LVQSGMILELQCLISQSHQVHETPRVAIMGSDTDLPMKDAAILDLFGVTHE  
VKIVSAHRTPEMYTATSAHSRGVQVITAGAGAAHLPGMVASITPLPVIGVPRAT  
RDGVDSLLSIVQMPRGVEVATVAINNATNAALLAVMLGISDTDLVSRMRQYQEDMR  
DENLANKGKLETFEGRESYLNO"  
complement(join(<66647. .66829,67056. .67229,67307. .67405,  
67776. .67976,68080. .68187,68273. .68651,68917. .69175,  
69620. .70064,70183. .>70242))  
/gene="F13M22.20"  
complement(<66647. .>70242)  
/gene="F13M22.20"  
complement(join(66647. .66829,67056. .67229,67307. .67405,  
67776. .67976,68080. .68187,68273. .68651,68917. .69175,  
69620. .70064,70183. .70242))  
/gene="F13M22.20"  
/codon\_start=1  
/product="CER1-like protein"  
/db\_xref="PID:g3236252"  
/translation="WASRPGFLTPWPTPLGSKFYLLAPLVFDSIYSYATIRDHKL  
LIVATVNRIVHSQWISLSYQTAKTKRIINKSIEFDQVDRDWDQIIFNTLIV  
YTKVYVSTGSIPEWRTDGVILVALLHAGPVEFYIYFPHRALHHFLYSRVHSHHS  
SVTEPITCATNSKPWLVIAVHPFAEHGITYLILGLPLITTCMGTVSVVSTALY  
LYIDPMNMGHCNELIPKLEFSLPLPKELCYTPSEHSLHHTQFRNYSLFEMMYD  
YIYGTDSDSLSEIYSEKEEKEDFADLHRLTSLDSIYHLRGLFASLSHSCSR  
YULFMKPAALISFLIRSFSTOTFVVRNRFRDLTLHSHLLPKFSHYMSHQKCEIN  
KNIEAILEADKVKGVMSLGNQGEELNGYEMVRRPKIKIRIVDGGSLAAEYV  
LHSIPGVGTLEFRQITKVARAIVFSLCQNAIKYMLVKKEHSLMAEFLDCKKENL  
VUTTNVYPMILVGDGLSTKEOKMAKDGLFLPFSQFPKTLRKDCFYHTTPAMILPH  
SAQNIDSCENWLGRRVMSAWRYGGIVHALEGKHECEGLDDNSIINPRVWEALRNG  
FQPLLLPSLET"  
complement(67680. .67715)  
/rpt\_family="AT\_rich"  
complement(70565. .70634)  
/note="exon predicted by xgrail, quality marginal"  
complement(70808. .73053)  
/gene="F13M22.21"  
complement(70808. .73053)  
/note="highly similar to GB:X95909(polyomorphism exists  
between nucleotides 73005 and 72085 which is a GA repeat.  
We found 6 copies in our sequence whereas only 5 copies  
exist in GB:X95909)"  
/gene="F13M22.21"  
complement(71000. .73027)  
/gene="F13M22.21"  
/codon\_start=1  
/product="receptor-like protein kinase"  
/db\_xref="PID:g3236253"  
/translation="MFLKLLTIFFFFFNLIQSSSQSLNFAYNNGFNPPDLSIOGI  
TTVTPNGLKLTNTVQKTHAFYTKPIRFKDSPNGTSSFSFVAIHSOAILSG  
HGIAFVAPNASLPYCNFSQYIGLFNANNGETNHVFAVELDTILSTFEDNTDNHV  
GIDINLSQVSSPAGYWDKQGFKNLTLSRKPMQVVDYDGRNTKIDVTWAPNEO  
KTPRLVTAVRLSSVLLQDMYVGFSSATGVSLSHYILGWSFGLNEKAPPLALSRLP  
KLUPREPKRISSEFYIGMPLISLFLIFSIFLVCYIVRRRRKFAEELEWEKEFGKNR

FRFKDLVYATKGFKEKGLLTGGGGSVYKGYMPGTFKLEIAVRKRVSHESROGKEFAVE  
IVSGRMHRNLVPLLGYCRRRGCELLLYDIMPNGSLDKYLYNTEPVTINWKQRIKVI  
LGVASGLFYLHEEWQVVIHRDVKASNVLLDGEELNGRLGDFGLARLYDHGSDPOQTHV  
VGTGLYAPETRTGTATMATDVFAGFAFLLEVACGRRPTEFQOQTDFTFLVDWVFG  
LWNGDIIAALKDPNMGSCDKEVEMVLKGLLCSHSDPRAPSRMQVLYHLRGDAKL  
LWNSPLDLSGGSMFGVHDGSFELGMSYSSSVYFKGTGSGSSTADSQLSGGR"  
72962. .73005  
/rpt\_family="(GAA)n"  
complement(73447. .73473)  
/rpt\_family="AT\_rich"  
join(<74148. .74206,74511. .74566,75952. .>77285)  
/gene="F13M22.22"  
<74148. .>77285  
/gene="F13M22.22"  
join(74148. .74206,74511. .74566,75952. .>77285)  
/gene="F13M22.22"  
/note="unknown protein"  
/codon\_start=1  
/db\_xref="PID:g3236254"  
/translation="MQIVPHVANAHMVKTSRKGALLALVVVGVECCPTRRCDLRPN  
LVDSNCULSRSTHOSLSVEDSIINHFRILNFRASIFLCRLGVICSTQSKMYGG  
KSYILKSVSLSIILITVTLTVLSAKNQPRISLIEVKTNLAKGEWVEDKKRPLYS  
GFECKOWLSNIFSCRMGRPDPSFEGYRWQPEGNCIPEFNRVNLRMRQNKTIATIGD  
SLGRFOFLCMATGKESPEVQNVSEYGLVIPKGAAPRCGHAYRPTTNTVLSY  
WSASLTDLVPHNNTDPPHLIAHMDRPFATIRYLHREHLVLTNGHWSRDKIERNH  
WVMHWTRVGGYEFKNENAKIFTIHSLVKWLDAQLPRLKCAFFTTISPRHEKCN  
NITPLSRGSKITGEGSLDTIVESAVNGTRYKILDTALSRLDEAHGACKLKPKKA  
SNVTSAPTNDCLHMLCPGIPGTWNELLIAQL"  
complement(75409. .75618)  
/rpt\_family="AT\_rich"  
complement(77388. .77424)  
/rpt\_family="AT\_rich"  
complement(77464. .77524)  
/rpt\_family="AT\_rich"  
77550. .77635  
/note="exon predicted by xgrail, quality  
excellent\_shadowexon"  
complement(join(<78252. .78416,78604. .79152,79555. .79945,  
80111. .>80604))  
/gene="F13M22.23"  
complement(<78252. .>80604)  
/note="predicted by genscan"  
/gene="F13M22.23"  
complement(join(78252. .78416,78604. .79152,79555. .79945,  
80111. .80604))  
/gene="F13M22.23"  
/note="hypothetical protein"  
/codon\_start=1  
/db\_xref="PID:g3236255"  
/translation="MAIFPFKLHHKPPSIPATTVPKPSHVLVAKVFFFTICIFISV  
AMISFYIFFGCDCHITGLRRGPENVLNTSSASSSTIRRNQSSSEADTDSHIA  
FGIGSIQTRDRDSRYSELWPNVTRGFIVLDEEPLNMTLWLSPPYQVSADTSRF  
SVTCWYGSRAIRMAIRIKETFELGLTVDRWFIMGDDTVFVDNLITVNLKYDNQM  
YVIGNSSEVEDIVHSYAMAYGGGIAISYPLAVELKLLDGCIDRASLYGSDQKI  
EACLSIEGVLTKELGPHQVDNPNGLAAHPVAPLTVLHLDIVDPIPFPTTID  
ALRRLVSAIKTDPDSRIQHSFCHDQTRNNYVSWSGYTIQIYTLTVLAKETPLTF  
KSWRTSSSEFSDTRPISDECPRLVYFLDRVYEVSGGTLTYTRKHVEGESTQC  
NSPDSRANPVEFIDVTTTLTPTDLWKMAPRRQCCEIVNSEEDSESVINVKIRHNPL  
ESVTPQSKHPNDEWTDVSLSTETI"  
complement(79324. .79361)  
/rpt\_family="AT\_rich"  
80698. .80727  
/rpt\_family="AT\_rich"  
complement(81151. .81260)  
/note="exon predicted by xgrail, quality excellent"  
82037. .82089  
/note="exon predicted by xgrail, quality  
excellent\_shadowexon"  
82508. .82558  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
complement(<83772. .>84686)  
/gene="F13M22.24"



e-mail: rounsley@tigr.org  
BAC clone F13M22 is from Arabidopsis chromosome II and is near the molecular marker ve018.  
The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tadb/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

## FEATURES

```

source
1..93845
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="ve018"
/clone="F13M22"
complement(10..29)
/notes="exon predicted by xgrail, quality marginal"
168..249
/notes="exon predicted by xgrail, quality good_shadowexon"
<464..>619
/genes="F13M22.1"
<464..>619
/notes="predicted by genscan"
/genes="F13M22.1"
<464..619
/genes="F13M22.1"
/notes="hypothetical protein, 5' partial"
/codon_start=1
/db_xref="PID:g3236260"
/translation="SRGDRVQRTHKDGDKRLDLPATRRKRLNATIASGTGFE
EIGNGL"
complement(join(835..1212,1373..1513,1600..1693,
1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
/genes="F13M22.2"
complement(835..>5654)
/genes="F13M22.2"
complement(join(1099..1212,1373..1513,1600..1693,
1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
/genes="F13M22.2"
/notes="unknown protein"
/codon_start=1
/db_xref="PID:g3236235"
/translation="MPKEENGOLKRDRLDDTDGNGKDHFPSPKQKAEASNDITSE
ISNPVSPVSTSLFRVDSQPKYKSLVEGSGDFGSEETVSDASVYGSQEQSSD
LSPSNVYKLETHLILGARVYKISTPPVQLLGIHSGYLCGCTCNFSKLSAYE
FEHAGAKTRPNHNLLENRAVINIYQELKTAAPVLEEVIRNVAAGSALNEGLRA
WKGDQGVNRNRYACLAGITLRIIPLSVNALLSVHVVLLSETDSCSCSYLGLGLDES
OSLTPCSVENHYSEKTYAKDTLDEPKRIAKLTSHVSGTGCHKKVSEGSNRKRDNL
HRLLFMPNGLPDGELAYVYKTIQSPSQFEAHGMAARRQPIRHIFISSGLSLHDIAM

```

```

SLANGHVITTTGDDMCICGDDGDLLLCAGCQPAFHATACLKFKQSPMEGTWYSSCND
GPSSKATTTDPSGNARPIVIRLSRVKAPESDIDGCVFCRSHDFSIGFDDRTVLL
CDOCEKYAVGCIKRENGEDLKEIPOEKWFCSCNSRIHTAVONSVSCGPQTLPTLL
DMICRDRKGIETDIDGVREILSGKSRYPHELPLLSRAAVIFRECFDPIVAKSGR
DLIPVYVIGNISQGFQGMICLVILVNSLVSAALLRIFGQVEALPLIVATISREYOG
RGYFQGLYACVENLLSLNVLNPAEAEASITWTKKFGFTKMSDQQLQEQYQEVOL
TIFKGTSMLEKVPKATGLSESTLI"
6010..6135
/rpt_family="(GGAA)"n"
6023..6202
/notes="exon predicted by xgrail, quality
marginal_shadowexon"
6179..6236
/rpt_family="POLY_A"
join(6647..6702,7161..7304,7322..>7499)
/genes="F13M22.3"
<6647..>7499
/notes="predicted by genscan"
/genes="F13M22.3"
join(6647..6702,7161..7304,7322..7499)
/genes="F13M22.3"
/notes="hypothetical protein"
/codon_start=1
/db_xref="PID:g3236236"
/translation="MTWKVREENKSFQFLFKFAWRPHFVYGATWTLTLTVAII
SLAPEAFVSAIFPSSSEVFSRRHIGSAGSLGLVTFGESEFEVEDGPRCAAGFCNS
CCGSRFRKRDYGPLGRRSPLI"
complement(join(<7761..7931,8019..8065,8154..8238,
8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
/genes="F13M22.4"
complement(<7761..>9886)
/genes="F13M22.4"
complement(join(7761..7931,8019..8065,8154..8238,
8323..8401,8494..8647,8824..8903,9001..9232,9769..9886))
/codon_start=1
/notes="putative ribitol dehydrogenase"
/db_xref="PID:g3236237"
/translation="MGIYVMTGKKGSGFSGASTAEDVTQATDASHLTAITGGTSG
IGLEAARVLAMRGHVIILAAARNKRAENKEMILQMNPNARVDYLQIDVSSISKVSRSF
VDQFLANVPLNLIINAGVMFCFKLTEDGIESQFATNHIGHFLTLNLLDKMKSTA
RESGVQIRVNLSSIAHTYVSEGIFQNGISQFATNHIGHFLTLNLLDKMKSTA
RLOEGVNITINSVHPGLVTTNLFYSGFSKMFVRAWTFLEKNIPOGATTCVVALH
PDLEGYTKYFGDCNIVAPSKFATNNSLADKLWDFSVFLIDSISK"
8638..8684
/rpt_family="AT-rich"
complement(9842..9915)
/rpt_family="(Ga)"n"
10503..10670
/notes="exon predicted by xgrail, quality
marginal_shadowexon"
10725..10746
/rpt_family="AT-rich"
complement(join(11416..11526,11608..12116,12551..12628,
12737..13604))
/genes="F13M22.5"
complement(11416..13604)
/genes="F13M22.5"
complement(join(11610..12116,12551..12628,12737..13522))
/genes="F13M22.5"
/codon_start=1
/product="putative ARF1 GTPase activating protein"
/db_xref="PID:g3236238"
/translation="MAAARLRLQSPENKVKVDCSKNPWASISYIGFMCLECSG
KHLGVLHVSFVSVMDSWEIQIKKMDAGNERNLNFLAQQISKETDIISKYNSN
AASVYDRIOALAEGRWDPPVIVKESVGGGLANKKPLSQGGGRSGENGWQND
DSFRDMMRNQAGDFRSRSGGAPAKSKASSEDIVSRQLEASAAKESFFAKRMAE
NESPGLPLPSQGGKYVGGFSGPAPRNQSGGGVDFVSMSEGRSLSLVAASAN
VVOQTMEFTSKYEGGLDQTVSETNVVASKTEIGQRTWIMKGYMAIASQVEEF
TKEASTNQONKTEGYQNSGIGNKTNANSFGSSSSSGHNSYRNSNSWDHG
EENNSKEAPKYSTNDDDDGGWAGWDDNDKDDDFYQYQASDKKSVGHNGKSDTAW
TGGGFL"
11697..11757

```

repeat\_region

misc\_feature /rpt\_family="(CAT)n"  
13716..13952  
/note="exon predicted by xgrail, quality excellent\_shadowexon"  
misc\_feature complement(14172..14309)  
/note="exon predicted by xgrail, quality marginal"  
mRNA join(14707..14911,15566..15765,15858..15964,16057..16189,16311..16602,16755..17294)  
/gene="F13M22.6"  
14707..17294  
/note="identical to GB:U40269"  
CDS join(15634..15765,15858..15964,16057..16189,16311..16602,16755..17182)  
/gene="F13M22.6"  
/codon\_start=1  
/product="origin recognition complex protein"  
/db\_xref="PID:g3236239"  
/translation="MEDINIEDEYGFERNYFLAKELGGAGKRSAHKLSDIHIVDEQ EURETASTIEMKHSKEISELMDYKIMYSKWFELRCFGLLMYFGSKKALVEDFAS ASLTDYVVVINGYLPVSNLQVLLALAEILLSKLLCKRKSGLSKGQETFPSRMD DILSFHGPSQGDQCFICVVVHNDIGPALRDPESQOTLARLSKSHIRLVASIDHVN APLLMDKQVHOFNWLWHVTEFAPYVGVGFPLVLAOGSTAQTAATAVLOSIT PQGNQVKETLAEVQLSHDPEDGMPDLDLYASASREFFVSSQVTLNSHLTFKDXHELVK TRNSDQGCLENIPLTSDAIRQLLDLQ"  
mRNA complement(<17532..>19016)  
/gene="F13M22.7"  
gene complement(<17532..>19016)  
/gene="F13M22.7"  
CDS complement(17532..19016)  
/gene="F13M22.7"  
/note="unknown protein"  
/codon\_start=1  
/db\_xref="PID:g3236240"  
/translation="WENHHPSTLLMSDSSASHEELDLEMMNNROSLLSGPPDINLPL SERSPPPPNLDACDILDVGLGQAVETENYMSVPVKGKCAKRVDSINGANEFF SYFPALEKSAKAIKIDNSINGDFKSDLDKLDVLFVHOHDMENIMWVFERPENAL SKYQORNGYNSRGQDLRFPEFSEKGFVSRMQKHVGLSNFCQVGHGIEPLPN LTCLEDEERKRWELTGRDINFIPPEASDFGSWRNLNPTDFELERPTPSLKNPTANH SKLLNGSLNSTOPSNSGEATDLSPPSHKRDLPFNGIHEECCLTVNPPQPV IEAHONELPTWNETGAMKNVYGPVTAKTIVIEDEEGLIILISLPFVDLNSVKYSWR NLTGIGIKVCLSTSRVPFIKRHRTFKLTDSESEHCPGFEVREIPLNSRIPEDAN IEAYDQGPSVLEILVPLKURAGEHEVRVCLRPNLGGNDMLT"  
misc\_feature complement(19045..19091)  
/note="exon predicted by xgrail, quality marginal"  
19455..19545  
misc\_feature /note="exon predicted by xgrail, quality excellent\_shadowexon"  
mRNA join(<20811..21507,21667..21805,22028..22434,22931..23006,23118..23233,23331..>23788)  
/gene="F13M22.8"  
gene <20811..>23788  
/gene="F13M22.8"  
CDS join(20811..21507,21667..21805,22028..22434,22931..23006,23118..23233,23331..>23788)  
/gene="F13M22.8"  
/codon\_start=1  
/product="putative zinc finger protein"  
/db\_xref="PID:g3236261"  
/translation="MFNNTTSEGGSGVIVVPTATVPTDTPCGTTTSTSTIIII GPPFPFAPPSIDITPLKLIFFVIAFAVPALVYALFENGPCSSNRSSSRSTSS SDTTPHATPTTETVTSEGGFKHDTSHKEIGNECSVCLMVFDTSDLRLOLSE CKAHFVLCIEWLKDPHCNCRIDTVSKVQOOTEAPNVNVRSGGNRRYSAT SRDDWRQGLPDASTVSVSRDVSQOAHPTVEDLILIGRLIHTPLPLTL SFCSRKPSDSSGLAPARNLAKTSIPREAYLVGTGDKGVKRLKALHAPRNYL LHLDEASDEERWELAYVRSEKKFENVMGLADLVTEKGTMLASTLHGVAILLK KAKDWDFTNLASDPLMPQDDILHIFSYLPRYLNFIETNIGMKENQARPIIID PGFYHLKSGVFAKERRSLPASFKLMSQTSVALT RPFLEFCIWMGDNLPTLRLMY TNFLLSSEGYFOTVVCNNKDYQNTVNHDLHYTKWDPQOORTLVNTVENFRDMVQGA PFARFERDDVLDKIDIELLGQTDGLTELEKTPDVVKPTVSKRLEKLMVRLLDHENE RAKQCK"  
repeat\_region complement(20872..20971)  
/rpt\_family="(TGG)n"

repeat\_region complement(21518..21543)  
/rpt\_family="AT-rich"  
misc\_feature complement(24200..24291)  
/note="exon predicted by xgrail, quality good\_shadowexon"  
repeat\_region complement(25078..25159)  
/rpt\_family="(TA)n"  
mRNA join(<23358..25408,25622..>26563)  
/gene="F13M22.9"  
23358..26563  
/gene="F13M22.9"  
CDS join(25358..25408,25622..26563)  
/gene="F13M22.9"  
/codon\_start=1  
/product="putative zinc finger protein"  
/db\_xref="PID:g3236241"  
/translation="WVFSIOAYLDSNNQOAPPNSNHDGTGASANGHVLRLPOLQP QOQOQPHNGSGGGGGGGSIRAGSMVDRARCANVALPEAALCKPCESTNTKFC YFNYSLQTPRHFCCTCRRYWTRGGALRNVPVGGCGRRNRRTKSNNNNNSTATSNN TSFSGNASTITSLSHYGGNESILSOILSPARKLMPTYNHLDGLTSTKTDNMS LNYGGLSQDLRSIHMASGGSLSMCSVDWRSASYHOOSMGGSNLESDSNPNPSANG LYSFSPRITTSASISSALASQFSSVKVEDNPYKWYNVANGCSSMNDLSAFGSSR"  
25691..25728  
/rpt\_family="(CAG)n"  
repeat\_region 26002..26054  
/rpt\_family="(CAA)n"  
repeat\_region complement(27038..27122)  
/note="exon predicted by xgrail, quality good\_shadowexon"  
28179..28208  
/rpt\_family="AT-rich"  
misc\_feature complement(28259..28328)  
/note="exon predicted by xgrail, quality marginal\_shadowexon"  
repeat\_region complement(28897..28980)  
/rpt\_family="(TA)n"  
repeat\_region 29085..29133  
/rpt\_family="AT-rich"  
misc\_feature complement(29945..30023)  
/note="exon predicted by xgrail, quality good"  
mRNA complement(join(<30476..30515,30659..30858,31168..>31269))  
/gene="F13M22.10"  
gene complement(<30476..>31269)  
/gene="F13M22.10"  
CDS complement(join(30476..30515,30659..30858,31168..31269))  
/gene="F13M22.10"  
/codon\_start=1  
/product="putative ribosomal protein L36"  
/db\_xref="PID:g3236242"  
/translation="MTTFAVTGLFVGLNKGHVTRRELAPRPNRSGKTSKRTFIR KLIREVAGMAPYEKRITELLKVGDKRALKVAKRKLGTGTHKRAKRREEMSSVLRKMS LGAAAAAEKKK"  
misc\_feature complement(31377..31401)  
/note="exon predicted by xgrail, quality marginal"  
misc\_feature complement(32006..32036)  
/note="exon predicted by xgrail, quality marginal"  
repeat\_region 32044..32161  
/rpt\_family="(GAAA)n"  
repeat\_region complement(32282..32354)  
/rpt\_family="(GAAAA)n"  
repeat\_region 32653..32718  
/rpt\_family="POLY-A"  
mRNA complement(join(<32935..33310,33630..>33739))  
/gene="F13M22.11"  
gene complement(<32935..>33739)  
/note="predicted by genscan"  
/gene="F13M22.11"  
CDS complement(join(32935..33310,33630..33739))  
/gene="F13M22.11"  
/note="hypothetical protein"  
/codon\_start=1  
/db\_xref="PID:g3236243"  
/translation="MEIGNRVADNVLNLREARHRRYFDYFIDEIWLFLGMSSET SLKEEVTIGGVKETREBECDDKLVDICLRKRLDEEEVEEVEENDGKFTPTRP ENRTPIVRECPAPMKRSSEMFGRMTYCRRLSLFLPEDDVNSFITDLQWRTTWTIK



gene  
complement(<53450. .>56602)  
/note="predicted by grail"  
/gene="F13M22.17"  
CDS  
complement(join(53450. .54129,54196. .54289,54371. .54453,  
54545. .54596,54711. .54868,54958. .56602))  
/gene="F13M22.17"  
/note="hypothetical protein"  
/codon\_start=1  
/db\_xref="PID:g3236249"  
/translation="MEDGNKLDPKKTMTMNEGLGDFEDDDRRPETHDRLLSSALT  
DMANASSDEDFDCRLSFSNAVSLTASRKFRTPAMSPDYDLWMAAPGISERR  
RLUHGMLASNDMWASVSRVSNAPVNGEKKMKMKNHVEVDQDDHVPVM  
LARSSESDIERLIEBKREEILGKISKORTYVSTICPTTRICQVOTPIRTPA  
VGNGALRGKGALVSMNSNAGFAFLIKNDITKEFIVNEYDDGWMNRLSDIQT  
CKGTLDEEKVCYGVSPVKELMRNENVRINVEPLMDKRNYSLSKVRSLKRGGA  
ALLNKIKGVAHNSLRVADKDVSDGSDSPKKGDKHGKANEMWVRVTRTGSKYKELS  
ALHMCQIQHAGAVNTIKFSQDAHLASGGADRVJHWEVECELSMNESGLPIIH  
PISCDSSGNETIVVERKKKGSSGRRHHNIPDYHVHVPETVFSFSDKPVCSLKGLHDA  
ILDLSWKSQLLSSMDKTVRLWDIETKCLKFAHNDYVTCIQSPVDENYFILSGS  
LDKARIWSTQDRHVWESDLHEMTAACYPDGOQALIGSHKIGICRAVDITDCKLSQ  
TNOIDVQNSKSOAKRKITTSFQSPVNPSEVLVTSADSRIRILDSSEVTHKFGFRNT  
CSGLSASQDCXYIKICASEDSQVYLKNDFHRTSLTILTSQHEHFCHQVSAAPWH  
GHVKGPPVQVHKSRHSKRISTSSQPSSTISPTKEEISATGPTISNRNKKSGLPPM  
PKAAATKSQIQPEEAPGELGSESEFSRSMNSEQHSRFGESPSINTSSLSWSWF  
DSGGHGPTQIQPTAWGVIVTATVHGSEIRSYQNFGLPFRIRGRQTTLF"  
complement(56145. .56194)  
/rpt\_family="(GAA)n"  
56886. .56744  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
complement(56743. .56808)  
/note="exon predicted by xgrail, quality marginal"  
complement(57092. .57128)  
/note="exon predicted by xgrail, quality marginal"  
57429. .57497  
/note="exon predicted by xgrail, quality  
marginal\_shadowexon"  
complement(57759. .57834)  
/note="exon predicted by xgrail, quality excellent"  
complement(join(<57887. .58388,59719. .59846,60078. .60132,  
60218. .60291,60384. .60527,61543. .>61581))  
/gene="F13M22.18"  
gene  
complement(<57887. .>61581)  
/note="predicted by genefinder"  
/gene="F13M22.18"  
CDS  
complement(join(57887. .58388,59719. .59846,60078. .60132,  
60218. .60291,60384. .60527,61543. .61581))  
/gene="F13M22.18"  
/note="hypothetical protein"  
/codon\_start=1  
/db\_xref="PID:g3236250"  
/translation="MPVRVVESTPAQASGTONVSNQOKEARWVNVQIQIDLEHG  
YLCGTMEALVPMADTPVITFWEGEIVDGKNYTFYTGKWEATREDDMRHWSKFPSP  
LQRMKQOYFNVGTDGLTIAGFYVCFSDGSISGFYDPNPSFHHMIISSKNVL  
KMEVEYSFKRKFTQSDDELSPKSKHTCFANVACSENTNGNSFTHDIYSNYSVN  
STTSMECNNDIEMKESGSGDKMISFESHDYIYGTQNLEDFSEKVIENILYLD  
QESEEEDAKGCSNAAKFVLLSSGRWTVNOGKENRILV"  
58770. .58795  
/rpt\_family="(GA)n"  
59165. .59234  
/rpt\_family="AT\_rich"  
59249. .59312  
/rpt\_family="AT\_rich"  
complement(59446. .59478)  
/rpt\_family="AT\_rich"  
complement(61713. .61851)  
/note="exon predicted by xgrail, quality good\_shadowexon"  
join(<62177. .62406,62814. .63035,63131. .63242,63337. .63399,  
63504. .63642,63940. .63995,64072. .64145,64247. .64394,  
64470. .64652,64757. .64844,64948. .65058,65257. .65354,  
65591. .65708,65788. .65882,66002. .66108,66213. .>66306)  
/gene="F13M22.19"

gene  
CDS  
complement(<62177. .>66306  
/gene="F13M22.19"  
join(62177. .62406,62814. .63035,63131. .63242,63337. .63399,  
63504. .63642,63940. .63995,64072. .64145,64247. .64394,  
64470. .64652,64757. .64844,64948. .65058,65257. .65354,  
65591. .65708,65788. .65882,66002. .66108,66213. .66306)  
/gene="F13M22.19"  
/codon\_start=1  
/product="putative phosphoribosylaminimidazole  
carboxylase"  
/db\_xref="PID:g3236251"  
/translation="MLLLKQSSAAVLVNGTTPVLHTSRSTYRVGPEFVTRTQSPQSL  
TMANQQAQLAKMILDPKNCSSASLSYGHMVDSDSDATVEEFKRGVGLGGQLG  
RMLCOASQALAKMILDPKNCSSASLSYGHMVDSDSDATVEEFKRGVGLGGQLG  
HYDVLTEKLEKQVDCQPKASTIRIIQDKYMKVHFQSHGIPLEPFMEISIEGAKK  
AGELFGYPLMKSKRLAYDGRGNVANQDELSSAVTALGGFSRGLYTEKWAFFVKEL  
AVIVARGDGMVCYPVETIHRDNICHIVKAPADVPWKINKLATDVAQKAVGSLEGA  
GVFAVELTDEDSOILLNEVAPRPHNSHQIECPTYSQFEQHLBNAVGLDLPDRSMR  
TPASIMILGEDDDVIDGEAGFKLAHRLIARALCIPGASVHWIDPEMKRQKMGHIT  
LVQSMGILEQRLOQLSEQSHVHETPRVAIIMGSDTDLPMKDAAILDLFGVTHE  
VKIVSAHRTPEMWTYATSAHSRGVQVIAGAGAAHLPGMVASITPLPVGPPVRAI  
RLDGVDLSLISVQMPRGVPVATVAINNATNAALLAVRLMIGISDTDLVSRMROYQEDMR  
DENLNGSKLIEGHWESYLNG"  
complement(join(<66647. .66829,67056. .67229,67307. .67405,  
67776. .67976,68080. .68187,68273. .68651,68917. .69175,  
69620. .70064,70183. .>70242))  
/gene="F13M22.20"  
complement(<66647. .>70242)  
/gene="F13M22.20"  
complement(join(66647. .66829,67056. .67229,67307. .67405,  
67776. .67976,68080. .68187,68273. .68651,68917. .69175,  
69620. .70064,70183. .70242))  
/gene="F13M22.20"  
/codon\_start=1  
/product="CER1-like protein"  
/db\_xref="PID:g3236252"  
/translation="NASRPGELTDPWPTPLGSKYLLAPLVDFDSIYSYATIRDHEKL  
LIVAVTVMRIVHSQIWSLSRYQAKTKRILKSIIEFDQVDRERTWDDQIIFNTLIV  
LTVKIVSGTSTIPETWRTDGVILVALLHAGPVEFYIYWFHRLAHHFILYSRVSHHS  
SVTPEITLCATNSKPWVLIIVAVVHPFAEHIGYTLILGLPLITTEMCTGFSVVSTALY  
LTYIDFMNMHGCNEFLPKLESLPLPKFLCYTPSPFSLHHTQFRNYSILFPMWYD  
YIYGTIDSCDSLYETSEKEERKPDALHRLTSLDLSIYHLRKFASLWSHPLSSRC  
YLFMKPFALISLILFRSFTQFVVRNFRDLTLHSHLLPKFSFSLWSHQKCCIN  
KNIEAAILKADKGVKVMGLNNGEELNGYGENYVRHPRKLRIRIVDGGSLAEVY  
LHSIPVGTKEVLFRGOITKVARAIVFSLCQNAIKVMRLRKEHSLABFLDCKENL  
VLTNNYPMIWLVGDLSTKEOKMAKDGTLFLPFSQFPKTLRKDCFYHTTTPAMLIPIH  
SQANDSCENLGRVMSANRVGGIVHALEKHECCGLDDNSIINPRVWEALRNG  
FQPLLLPSET"  
complement(67680. .67715)  
/rpt\_family="AT\_rich"  
complement(70565. .70634)  
/note="exon predicted by xgrail, quality marginal"  
complement(70808. .73053)  
/gene="F13M22.21"  
complement(70808. .73053)  
/note="highly similar to GB:X95909(polymorphism exists  
between nucleotides 73005 and 72085 which is a GA repeat.  
We found 6 copies in our sequence whereas only 5 copies  
exist in GB:X95909)"  
/gene="F13M22.21"  
complement(71000. .73027)  
/gene="F13M22.21"  
/codon\_start=1  
/product="receptor-like protein kinase"  
/db\_xref="PID:g3236253"  
/translation="MFLKLLTIFFFFFNLIFQSSQSQSLNFAYNNGFNPTDLSLTOGI  
TTVTNGLKLNTTVQTKGHAFYTKPIRKDSPTNGTSVSSFSVFALHSQIALUSG  
HGIAFYFVNASLPYGNPSQYIKFLNANNGETNHFVAVELDTILSTFENIDNNHV  
GDINSLKVSQSPAGYWDKQFNKTLISRKPMQVWYVDYDRTNKIDVTWAPENDY  
KPTRPLVAVRDLSSVLLQDMVKVGSSTGSLSEHYILGWSEGLNEKAPPLASRLP  
KLPRPEKPISEFYKIGHPLISLELFIPLVCYIVRRRFAKEFELEWEKFEFGNRL  
FRFKDLYATKGFKGKGLGTGTFGSVYGVMPGKLEIARVXVSHESQGWKEFVAE  
IVSIGRMSHRNLVPLLYGCRRRRGLLYDYMPNGSLDKLYLNTPEVTLNWKORIKVI

```
LGVASGLFYLHEEQVVIHROVKASNVLLDGLGNLGRDGLGLARLYDHGSDPQTHV
VCTGLYLAPERTGTGRMATDVFAGFAGLELVACGRRIEFQOETDEFLLVDVFG
LNWKGDLAAGPNMGSCDEKEVENYKLGLLCSHSDPRARPSQVHLVLRGDAKL
PELSPLDVLGSGMMFGVHDGFSLEGMSSYSSSVFKGFTGSSSTADSQLSGGR"
72962..73005
/rpt_family="(GAA)n"
/complement(73447..73473)
/rpt_family="AT_rich"
join(<74148..74206,74511..74566,75952..>77285)
/gene="F13M22.22"
<74148..>77285
/gene="F13M22.22"
join(74148..74206,74511..74566,75952..77285)
/gene="F13M22.22"
/note="unknown protein"
/codon_start=1
/db_xref="PID:g3236254"
/translation="WQIVPHVANAHMVKTSRKGALLVVGVECCPTRRCDLRDPN
LVVDSNCLLSRSTHQSLLVEDSIIINHRFILNFRASIFLCIRIGVICSTOSKMVG
KYILLKGSVLSLIIILLVTLVLLYSAKNQPRIEYKTCNLAKGEWVEDKKRPLYS
GFECKWLNIFSCRVMGRPDFSEFGRWQPGECNIEFNRRVFLRMQNKTIATIGD
SIGRFOFLMCATMGKESPEVQNVGSEYGLVPGKAPPGGWAYREPTTNTVLSY
WSAFTDLVMMNTDPHLIAMHLDPRPAIRNYLHREHLVLTGHHWSRDKIKNH
WYMHVNGTREVGGYFKNVENAKIFTIHSLVKWLDAQPLHPRLKAFFTISRHRKCN
NTPLSRSGKITGEGSLDTIVESAVNGTRVKILDITALSKLDRHAHAGCKLKPKKA
SNVTSAPTFNDGLHCLGPIPTWNELLIAQL"
complement(75409..75618)
/rpt_family="AT_rich"
complement(77388..77424)
/rpt_family="AT_rich"
complement(77464..77524)
/rpt_family="AT_rich"
77550..77635
/note="exon predicted by xgrail, quality
excellent_shadowexon"
complement(join(<78252..78416,78604..79152,79555..79945,
80111..>80604))
/gene="F13M22.23"
complement(<78252..>80604)
/note="predicted by genscan"
/gene="F13M22.23"
complement(join(78252..78416,78604..79152,79555..79945,
80111..80604))
/gene="F13M22.23"
/note="hypothetical protein"
/codon_start=1
/db_xref="PID:g3236255"
/translation="MAIFDPFKLHKPPSIFPATVVKPSHVLVLVAKFFFTICIRISV
AMISFIIIFGSDCHHTIGLRLGPENVNLTNSASSSTIRRNQSEADISHIA
FGIGSIQTRDRSRSELWWRPNVTFIWLDEEPLNMTWLSTSPPYVSADTSRF
SYTCWGSRAIRMARIIKETELGLTDVWFIMGDDDTVFVDNLITLVNKYDHNQM
YVIGNSESVEODIVHSYAMAYGGGIAISYPLAVELVLLDGCIDRVSALYSGDKOI
EACLSFIVPLAKELGFHOVDIRGNPYGLAAHPVAPVLTVHLHDVDPDFPFTQID
ALRRLVSAYKTPDSRIHQSFCHDQTRNNYVSVSGYTIQIPTLVTAKELETPTLF
KMRUTSSSPFSFTRIPDSFCDPCERPLVFLDRVYEVGSQLTITTRKRVGEVSTQC
NSPDYRANPVEFIDVSTTLTLPDLKMKAPRQCCIEVNSEDSSESVINVKIRHFNPL
ESVTPQSKHPNDEWTSVLSTETI"
complement(79324..79361)
/rpt_family="AT_rich"
80698..80727
/rpt_family="AT_rich"
complement(81151..81260)
/note="exon predicted by xgrail, quality excellent"
82037..82089
/note="exon predicted by xgrail, quality
excellent_shadowexon"
82508..82558
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(<83772..>84686)
/gene="F13M22.24"
complement(<83772..>84686)
/gene="F13M22.24"
```

```
CDS
complement(83772..84686)
/gene="F13M22.24"
/codon_start=1
/product="putative zinc finger protein"
/db_xref="PID:g3236256"
/translation="MEKPGFWIPKKSNKSEWEELAFEDDAAGSLWPPRSYTCGFC
RREFKSAQALGCHMVHRRDRARLKQADDQYLFKPKSSSPSEYPKSHKSDNIHETCYT
LVFNTPKNVFKTOHSCVIDLSSSLPYLTPSRVSSGLPGKQHTSSSPSPFVVEPSKN
SKYIPSSSPWSCPSVIVQKSCDLYEIPAMGEKARKTESDVPKIGHRAKLSLGNWTD
LSVSNLVIHQSFPITAHGSDDEEIGRVDIRKRRRHESPQOOSIFISSLSCKSDIITR
NEESKHGDRFEDLDLELRLGTDPKGI"
84230..84321
/rpt_family="(GAA)n"
84910..84962
/rpt_family="AT_rich"
85196..85220
/rpt_family="(TA)n"
complement(85246..85307)
/rpt_family="AT_rich"
complement(85742..85810)
/note="exon predicted by xgrail, quality marginal"
86102..86122
/rpt_family="AT_rich"
86195..86308
/note="exon predicted by xgrail, quality good_shadowexon"
86660..86687
/rpt_family="AT_rich"
86972..87002
/note="exon predicted by xgrail, quality marginal"
87049..87569
/gene="F13M22.25"
87049..87569
/gene="F13M22.25"
87120..87332
/gene="F13M22.25"
/note="unknown protein"
/codon_start=1
/db_xref="PID:g3236257"
/translation="MQIRYGNESQQLSSYYVRLPGDSGRFVRSDLDSVSGPSNSNR
ASSSTSTTFAVYAGVQSPVSRQVVT"
join(<88062..88169,88258..88308,88608..88684,88770..88857,
89934..89209,89301..>89585)
/gene="F13M22.26"
<88062..>89585
/gene="F13M22.26"
join(88062..88169,88258..88308,88608..88684,88770..88857,
89934..89209,89301..89585)
/gene="F13M22.26"
/codon_start=1
/product="putative alcohol dehydrogenase"
/db_xref="PID:g3236258"
/translation="MAAPTRFPELNTGAKLPCVGLGTAVMATAEQAIKIGVRHDC
ASIVNEKEIGVGLKLIGDGFVKREELFITSKLNSNDHLPEDVPKALEKTIQDLQID
YVDLLIHWHPASLKRESLMPTEMLTKPDITSTWKAMEALYSGKARALGVSNFSSKK
LTDLLNARVTPAVNQCHPVVQQQGLHCLKSGVHLUSGYSPGSGKGEVRLKVL
QNPIYVEAEKLGKTTAQVALRWGLQTHSVLPKSSSGARLKENLDVDFWSDPEDLFT
KFSNIPOASSFFLFFSR"
89723..89798
/note="exon predicted by xgrail, quality excellent"
complement(89743..89842)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
complement(90572..90624)
/rpt_family="AT_rich"
join(<90955..91074,91214..91264,91485..91561,91649..91736,
91942..92217,92291..>92551)
/gene="F13M22.27"
<90955..>92551
/gene="F13M22.27"
join(90955..91074,91214..91264,91485..91561,91649..91736,
91942..92217,92291..92551)
/gene="F13M22.27"
/codon_start=1
```















[illegible]

```

BASE COUNT      5345 a   1560 c   1507 g   4256 t
ORIGIN          TLLIPLITTEVRNRSVVKRIYDLLLLLEKRRFGNNLD"

Query Match           7.2%    Score 44.2: DB 3; Length 12668;
Best Local Similarity 55.6%; Pred. No. 1.8;
Matches 89; Conservative 0; Mismatches 68; Indels 0; Gaps 0:

Qy 10 atggatgagaacaggtttcaacagttaagtcttcaaaacaacaagaacatcatggcgaagacag 69
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9280 ATTCATCAAAATCACAACAACAACAAATTAGAACACAACAACAACAACAACAACAACAAC 9339

Qy 70 atggagtcacttgataaacacaattggcgacaataaacacaacaataatgtttacaagctca 129
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9340 AACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 9399

Qy 130 ttacgagataattatggaaatgaagatcatta 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9400 AAACCTCATTAATTCATCATTCATCATCAACTA 9432

```

	Query Match	7.2%	Score 44.2	DB 3	Length 12668
	Best Local Similarity	55.6%	Pred. No. 1.8		
	Matches 85	Conservative 0	Mismatches 68	Indels 0	Gaps 0
Qy	10 atggtatgaacacgatttcaacagtaagttaacttcaaaacaaaagcatcatggcgaagacag	69			
Db	9280 ATTGATCAATTCACACACACACACACANTTACAAACAAACAAACAAACAAACAAAC	9339			
Qy	70 atggagtgacttgtaatacaacaaatggcgacaataaacacacaaataatgtttacaagctca	129			
Db	9340 AACAAACAACAACAACAACAACAACAACAACAACAACAACAATCAATCAATCAAC	9399			
Qy	130 ttacagatataattatggaaatgaagatcattta	162			
Db	9400 AAATCTGATTAAATTCATCATCATCATCACTA	9432			

Search completed: July 15, 1999, 13:56:12  
Job time: 3934 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 13:58:17 ; Search time 62 Seconds  
(without alignments)  
1875.243 Million cell updates/sec

Title: US-09-156-580-1\_COPY\_190\_807  
Perfect score: 618  
Sequence: 1 atggcaggcatgataagaaa.....aacttcgactggatccact 618

Scoring table: IDENTITY\_NUC

Searched: 240622 seqs, 94065609 residues

Database : N\_Geneseq\_34:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	40.8	6.6	198	1	TL17202	DNA-spanner oligon
2	40.8	6.6	198	1	TL17203	DNA-spanner oligon
3	40.6	6.6	1890	1	Q92524	Nicotiana glauca ar
4	39.2	6.3	144	1	Q12515	CSP-2 peptide from
5	39.2	6.3	198	1	TL17204	DNA-spanner oligon
6	39.2	6.3	198	1	TL17205	DNA-spanner oligon
7	37.6	6.1	120	1	Q12516	CSP-2 peptide from
8	35.6	5.8	117	1	Q12514	CSP-2 peptide from
9	35.4	5.7	9636	1	Q67190	P. falciparum tran
10	35.4	5.7	1942	1	TL38896	Carnation ACC synt
11	35	5.7	2160	1	TL85328	Truncated murine T
12	34.8	5.6	1016	1	N93636	Sequence of alpha-
13	33.8	5.5	8920	1	Q62924	Carbamoyl-phosphat
14	33.8	5.5	58407	1	V21210	Methanococcus jann
15	33.6	5.4	780	1	V26001	Bacteriophage DNA
16	33.6	5.4	2519	1	V54124	Human membrane pro
17	33.4	5.4	110000	1	T58840_2	Continuation (3 of
18	33.4	5.4	110000	1	T58840_3	Continuation (4 of
19	33.2	5.4	3600	1	T77330	Solanum tuberosum
20	33.2	5.4	1813	1	V34193	Human secreted pro
21	33	5.3	9295	1	T62125	Arabidopsis thalia
22	33	5.3	5318	1	V20701	Cryptosporidium pa
23	33	5.3	5163	1	V20700	Cryptosporidium pa
24	32.8	5.3	3101	1	Q02047	Sequence encoding
25	32.8	5.3	1430	1	Q92525	Nicotiana plumbagi
26	32.8	5.3	1040	1	Q92530	P. communis (pear)
27	32.6	5.3	2183	1	Q54769	T. parva sporozoit
28	32.6	5.3	15229	1	V18276	RSV isolate 18537
29	32.4	5.2	4467	1	T68648	PSRQ800 fragment 1
30	32.4	5.2	15225	1	V17552	Respiratory syncyt
31	32.4	5.2	15219	1	V18277	RSV vaccine 2B33F
32	32.4	5.2	15219	1	V18278	RSV vaccine 2B20L
33	32.4	5.2	15219	1	V18279	RSV revertant 2B33
34	32.4	5.2	15219	1	V18280	RSV revertant 2B20
35	32.4	5.2	15218	1	V18275	RSV isolate 2B wil
36	32.4	5.2	110000	1	V21209_15	Continuation (16 o
37	32.2	5.2	1055	1	Q94335	Degenerate Alterom
38	32	5.2	7241	1	Q15140	Genomic clone LE-A
39	32	5.2	6106	1	Q74678	Bacillus cereus ve
40	32	5.2	6049	1	TL19339	B. cereus VIP2A(a)
41	32	5.2	2417	1	T51537	Mycobacterium gall
42	32	5.2	350	1	T65080	Canine genomic mic
43	32	5.2	6049	1	T73993	B. cereus VIP2A(a)

C 44 32 5.2 6049 1 V16165 DNA encoding veget  
45 32 5.2 7244 1 V15705 Tomato ACC synthas

ALIGNMENTS

```
RESULT 1
TL17202
ID TL17202 standard; DNA; 198 BP.
AC TL17202;
DE DNA-spanner oligonucleotide BAM36-198.
KW DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW complementary sequence; anti-test protein antibody; extension;
KW primer template complex; double stranded duplex; amino acid;
KW sequence determination; epitope ordering; restriction mapping; ss.
OS Synthetic.
PN WO9605847-A1.
PD 29-FEB-1996.
PF 22-AUG-1995; U10668.
PR 22-AUG-1994; US-294133.
PA (UYPE-) UNIV PENNSYLVANIA.
PI Eberwine J.
DR WPI: 96-151138/15.
PT Sequencing proteins by epitope ordering and protein restriction
PT mapping - allows for characterisation of post-translational events
PT on the protein
PS Example 3; Page 18; 28pp; English.
CC The DNA-spanner oligonucleotides BAM-36-198, BAM36-198op, RI36-198
CC and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC restriction site, and a long nucleotide sequence contg. a sequence
CC at the 3'-end capable of annealing to a complementary sequence,
CC i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC RI36-198op, respectively. Binding a complementary pair of oligos to
CC an anti-test protein antibody (Ab), which is then bound to the
CC protein, allows the oligo sequences to form a primer template
CC complex, which can be extended to form a double stranded duplex.
CC The duplex can be digested with a restriction enzyme, the prods.
CC separated and their band pattern detected, allowing a portion of
CC the protein's amino acid sequence to be determined.
CC This method allows the determination of a protein sequence by
CC epitope ordering with single, di- and tri-amino acid specific Abs,
CC followed by restriction mapping. The procedure provides enough
CC protein sequencing to allow library screening oligo probes to be
CC generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC the epitope ordering procedure allows post-translational
CC modification events to be characterised.
SQ Sequence 198 BP; 121 A; 64 C; 7 G; 6 T;

Query Match 6.6%; Score 40.8; DB 1; Length 198;
Best Local Similarity 61.1%; Pred. No. 0.038;
Matches 66; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 9 catggatagaacacagttctcaacagtaagtaactctcaaaaacaaagcattcgagaca 68
DB 34 CRAGGATGCCAACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 93
QY 69 gatggagtactgataatacaacaatggcgacaataacaacaataaa 116
DB 94 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 141

RESULT 2
TL17203
ID TL17203 standard; DNA; 198 BP.
AC TL17203;
DE DNA-spanner oligonucleotide BAM36-198op.
KW DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW complementary sequence; anti-test protein antibody; extension;
KW primer template complex; double stranded duplex; amino acid;
```







CC cross-reactive with P. berghei antiserum. The Pf and Pb  
CC proteins are substantially homologous, if not identical.  
SQ Sequence 120 BP; 70 A; 40 C; 10 G; 0 U;

Query Match	6.1%	Score 37.6;	DB 1;	Length 120;
Best Local Similarity	59.3%	Pred. No. 0.22;		
Matches 64;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

19	Qy	acacgattcaacagtaagctacttccaaacacaaagaacatcatggcaagacagatggagtac	78
1	Db	AGCAGCAACCAACCAACCAACAGCAGCAACCAACAGCAACCAACCAACCAACCAAC	60
79	Qy	ttrgaatacaucaatggcgacataaacaacaacaataatgttacaagc	126
61	Db	AACACCAACCAACCAACAGCAGCAACCAACCAACCAACCAACCAACCAACCAAC	108

## RESULT 8

Q12514 ID Q12514 standard; DNA; 117 BP.  
AC Q12514;  
AD 18-SEP-1991 (first entry)  
DE CSP-2 peptide from P. falciparum (clone 3).  
DE P. berghei; Pf42; Pf54; Pb42; Pb54; ss.  
KW Plasmodium falciparum.  
OS WO9108756-A.  
PN 27-JUN-1991.  
PD 12-DEC-1990; U07194.  
PF 12-DEC-1989; US-448109.  
PR (BIOM-) BIOMEDICAL RES INST.  
PA (BIOM-) BIOMEDICAL RES INST.  
PI Hollingdal MR, Sina B;  
PFI 91-207865/28.  
P-PSDB; R12902.  
DR Sporozoite polypeptide antigens - are useful for antimalarial  
PT vaccines and the encoding these peptide(s) can be incorporated  
PT into virus to produce live vaccines  
PS Disclosure: Page 21; 32pp; English.  
CC A portion of the gene encoding the P. falciparum sporozoite  
CC antigenic protein CSP-2 (mol. wt. 42kD/54kD) was cloned. Clones 3,  
CC 4 and 6 were identified and found to contain malaria DNA inserts as  
CC represented in Q12514. Q12515 and Q12516 respectively. The  
CC polypeptides are present on the surface of the sporozoites and are  
CC cross-reactive with P. berghei antiserum. The Pf and Pb  
CC proteins are substantially homologous, if not identical.  
SQ Sequence 117 BP; 69 A; 39 C; 9 G; 0 U.

Query Match	5.8%	Score 35.6;	DB 1;	Length 117;
Best Local Similarity	60.2%;	Pred. No. 0.72;		
Matches 59;	Conservative	0;	Mismatches 39;	Indels 0;
				Gaps 0;

19	Qy	aacagtttcaacagtaagctacttcaaaaaaacaagaagcatcatgtgcgaagacagatggagtac	78
1	Db	AGCAGCAACACACACACACAGCAGCAGCAACAGCAGCAACAGCAGCAACACACACACAC	60
79	Qy	ttagtaacaacaatggcgacaataaacaacaacaataa	116
61	Db	AACACACACACACACACAGCAGCAACACACACACACACACAA	98

## RESULT 9

Q67190	Q67190 standard; DNA; 9636 BP.
ID	Q67190.
AC	20-FEB-1995 (first entry)
DE	P. falciparum transmission blocking target antigen Pf230 DNA.
DT	Protozoan; transmission blocking target antigen; Pf230; malaria;
KW	vaccine; ss.
OS	Plasmodium falciparum.
OS	Location/Qualifiers
FH	149..9553
FT	/*tag= a
FT	

PN	W09417187-A.
PP	04-AUG-1994.
PD	18-JAN-1994; U00547.
PE	29-JAN-1993; US-010409.
PF	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
PG	(USSH ) US SEC DEPT HEALTH.
PH	Kaslow DC, Williamson KC;
PI	WPI: 94-264101/32.
PJ	P-PSDB; R57474.
PK	New Plasmodium falciparum transmission blocking target antigen -
PL	useful in antimalarial vaccines, also related DNA, expression
PM	vectors and transformed cells
PN	Claim 2; Page 24; 63pp: English.
PO	The DNA may be used to express Pf320 in a host cell. It can also
PP	be used in vaccines (by incorporation into viral vectors which are
PQ	then used to infect host cells) and oligonucleotides derived from it
PR	can be used to identify homologous proteins in other spp.
PS	Sequence 9636 BP: 4156 A: 974 C: 3084 T:
PT	

```

Query Match          5.7%; Score 35.4; DB 1; Length 9636;
Best Local Similarity 46.8%; Pred. No. 3;
Matches 111; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```

Qy	82	aataacaacaatg	cgacaaataa	caacaacaataa	tgttacaagctc	cattacgagataa	141
Db	8177	AATCAAAACAAT	GTGCTATTAG	TAAATAAATAAT	TATCTTATATTTACACAT	8236	
Qy	142	tatggaaatgaag	atcatttactt	ggfggactat	ctctctggcct	ccaagatcttataca	201
Db	8237	ANTAAATAATC	AGAGAATAAT	CTAATATGTG	TATATCTTTTAA	TTCCAAAACCTGTATA	8296
Qy	202	tgtagctttt	gtaaaaaggg	atttagatc	tgctcaagctct	tggtggacacatgaagtt	261
Db	8297	GGATTAAT	ATGTCCTA	TATAAAAA	TTAATCCCAACAC	TGTTTTGTGAGTGTATTAT	8356
Qy	262	catagaagagat	agagccattt	gtgagacaat	ccaccctctagagata	ttaataggtat	318
Db	8357	GTTAAACAAGA	AGATGTAC	CTTTCGAAAC	TATACACGCTGAT	ATAATATAATACATTT	8413

RESULT 10

T38896	
ID	T38896 standard; DNA; 1942 BP.
AC	T38896;
DT	10-FEB-1997 (first entry)
DE	Carnation ACC synthase gene.
KW	Carnation; l-aminocyclopentane-l-carboxylic acid synthase;
KW	ACC synthase; co-suppression; ethylene; senescence;
KW	transgenic plant; ss.
OS	Dianthus sp. cv. White Sim.
FH	Location/Qualifiers
FT	134..1687
FT	cds
FT	/*tag= a
FT	complement (287..308)
FT	/*tag= b
FT	/note= "5' primer for truncated ACC oxidase gene"
FT	1352..1373
FT	primer_bind
FT	/tag= c
FT	/note= "3' primer for truncated ACC oxidase gene"
PN	WO9635792-A1.

WO9635792-A1.

PD 14-NOV-1996.  
 PF 09-MAY-1996; AU0286.  
 PR 09-MAY-1995; AU-002862.  
 PA (ALLR-) ALLRAD NO 1 PTY LTD.  
 PA (FLOR-) FLORIGENE INVESTMENTS PTY LTD.  
 PI Cornish EC, Graham MW, Gutterson NI, Michael MZ;  
 PI Tucker WT;  
 DR WFI; 96-518680/51.  
 DR P-PSDB; W04558.  
 DR Producing transgenic plants, with reduced climacteric ethylene  
 PT prodn. - to give flowers and buds, specifically carnations, with  
 PT increased post-harvest life

Db	201	TATGAACAAAAGAAATACGCGAGCAACAGGAACAGCTGTGTCATTTTCAACAAAAACCTTAATTGAC	260
Qy	123	aagctca	129
Db	261	AGGCTGA	267
RESULT 12			
N93636	ID	N93636 standard; DNA; 1016 BP.	
AC	N93636;		
DE	22-MAR-1991	(first entry)	
DE	Sequence of alpha-gliadin gene fragment used to construct plasmid		
DE	pxy32		
KW	Celiac disease diagnosis; dough formation; food technology; ds.		
OS	wheat.		
PN	US4826765-A.		
PD	02-MAY-1989.		
PF	24-OCT-1986; 922616.		
PR	24-OCT-1986; US-922616.		
PA	(UWHA-) Univ of Hawaii (USDA).		
PI	Greene FC, Stiles JI, Neill JD, Anderson OD, Litts JC;		
DR	WPI; 89-150270/20.		
PT	New wheat gluten protein gene plasmid - useful for genetic		
PT	transformation of yeasts to produce gliadin or glutenin		
PS	Example; Fig 5; 21pp; English.		
CC	The preferred gene fragment in the patent of the invention encodes		
CC	alpha-gliadin or glutenin, and the preferred plasmid encoding alpha-		
CC	gliadin is PAY 31. The plasmid may be used for the prodn. of alpha-		
CC	gliadin, which may be used in the food technology industry to modify		
CC	prodn. processes of doughs and batters and to reduce unit costs.		
CC	Gliadin may also be used for diagnosis and treatment of illness		
CC	caused by wheat gluten proteins and for testing theories of dough		
CC	formation.		
SQ	Sequence. 1016 BP; 359 A; 321 C; 141 G; 195 T;		
Query Match 5.6%; Score 34.8; DB 1; Length 1016;			
Best Local Similarity 52.8%; Pred. No. 2.2;			
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps			
Qy	41	tcaaaacaaagcatcatgagcaagacagatggagtacttgataacaacaatggcgaca	100
Db	345	TCTACAAACCAACCACTTTTCGCAGCAACAGCAACAACAACAACAACAACA	404
Qy	101	ataacaacaacaataaagtgtacaaagctcatcagagataattatgaaatgaagatcatt	150
Db	405	ACAACAACAACAACAACAACAACAACAACAACAAATCTTCAACAATTTTGCACAACAAC	464
Qy	161	tacttggtgactattctcttg	182
Db	465	TCCATGCAGGGAGTTGTCCTTG	486
RESULT 13			
Q62924	ID	Q62924 standard; cDNA; 8920 BP.	
AC	Q62924;		
DR	06-DEC-1994	(first entry)	
DE	Carbamoyl-phosphate-synthetase II.		
KW	Carbamoyl-1-phosphate-synthetase II; CPSII; pscPSII gene;		

OS	Plasmodium falciparum.
FH	Key Location/Qualifiers
FT	1226 .8401
FT	cds /*tag= a
FT	/EC_number= 6.3.5.5
PN	W09412643-A.
PD	09-JUN-1994.
PF	02-DEC-1993; AU0617.
PR	03-DEC-1992; AU-006206.
PR	16-DEC-1992; AU-006380.
PA	(UNIX ) UNISEARCH LTD.



Qy 273 tagagccattttagagacaatcaccacccctagagatat 308  
||| ||||| | ||||| |||||  
Db 246 TAACTTAACTTTTGGCTTATATAAACCACTAAAGATAT 281

Search completed: July 15, 1999, 13:58:23  
Job time: 3939 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 12:07:21 : Search time 377.02 Seconds  
(without alignments)  
2500.430 Million cell updates/sec

Title: US-09-156-580-1\_copy\_190\_807

Perfect score: 618

Sequence: 1 atggcaggcatgatgagaa.....aacttcgactgggattccact 618

Scoring table: IDENTITY\_NUC

Searched: 2002476 seqs, 762712212 residues

Database :

EST:\*

1: em\_est1:\*

2: em\_est2:\*

3: em\_est3:\*

4: em\_est4:\*

5: em\_est5:\*

6: em\_est6:\*

7: em\_est7:\*

8: em\_est8:\*

9: em\_est9:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: gb\_est6:\*

16: gb\_est7:\*

17: gb\_est8:\*

18: gb\_est9:\*

19: gb\_est10:\*

20: gb\_est11:\*

21: gb\_est12:\*

22: gb\_est13:\*

23: gb\_est14:\*

24: gb\_est15:\*

25: gb\_est16:\*

26: gb\_est17:\*

27: gb\_est18:\*

28: gb\_est19:\*

29: gb\_est20:\*

30: gb\_est21:\*

31: em\_est10:\*

32: em\_est11:\*

33: em\_est12:\*

34: em\_est13:\*

35: em\_est14:\*

36: em\_est15:\*

37: em\_est16:\*

38: em\_est17:\*

39: em\_est18:\*

40: em\_est19:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.4	6.9	303	31 AU034213	Dictyostele
2	42.2	6.8	482	28 C91001	Dictyosteli

	3	41.8	6.8	419	31	AU033391
c	4	41.4	6.7	358	27	AA634656
	5	41.4	6.7	591	31	AU033325
c	6	40.6	6.6	554	19	C23753
	7	39.8	6.4	574	28	C91220
	8	39.4	6.4	312	23	AI163084
	9	39.4	6.4	397	31	AU034056
	10	39.2	6.3	472	12	N39467
	11	39.2	6.3	552	28	C91403
	12	39.2	6.3	480	28	C93684
c	13	39	6.3	402	11	R99317
	14	39	6.3	546	17	C22974
	15	39	6.3	246	31	AU033454
	16	39	6.3	617	31	AU033655
	17	38.8	6.3	198	31	AU034673
	18	38.6	6.2	466	20	C25660
	19	38.4	6.2	408	31	AU033476
	20	38.2	6.2	249	28	C94333
	21	38.2	6.2	526	31	AU033912
	22	38	6.1	627	28	C91439
	23	38	6.1	294	31	AU034848
	24	37.6	6.1	585	20	C24619
	25	37.6	6.1	606	25	C84091
	26	37.6	6.1	427	28	C91312
	27	37.6	6.1	287	28	C94188
	28	37.4	6.1	539	29	AI070145
	29	37.4	6.1	414	31	AU033929
	30	37	6.0	470	26	C84849
	31	36.8	6.0	545	25	C83861
	32	36.6	5.9	503	29	AI069527
c	33	36.2	5.9	422	16	AA125493
	34	36.2	5.9	553	22	AA528418
	35	36.2	5.9	458	23	AI291213
c	36	36.2	5.9	630	26	AA744771
	37	36.2	5.9	526	26	AA775720
c	38	36	5.8	528	28	C92028
	39	36	5.8	531	28	C93163
	40	35.8	5.8	486	20	AA518458
	41	35.8	5.8	533	25	AA710415
	42	35.8	5.8	278	30	AI153808
	43	35.6	5.8	524	28	C91295
	44	35.4	5.7	484	20	C25586
	45	35	5.7	456	10	R53852

#### ALIGNMENTS

RESULT 1

AU034213

LOCUS

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

AU034213 303 bp mRNA EST 28-OCT-1998  
Dictyosteleium discoideum siug cDNA, clone SLC264, mRNA sequence.

AU034213

g3799637

EST

Dictyosteleium discoideum (strain:AX4) slug cDNA to mRNA,

clone:lib:SL.

Dictyosteleium discoideum

Eukaryota; Dictyosteliida; Dictyosteleium.

1 (bases 1 to 303)

Urushihara, H.

Developmental cDNA in Dictyosteleium discoideum

Published Only in Database (1998) In press

2 (bases 1 to 303)

Urushihara, H.

Direct Submission

Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko

Urushihara, University of Tsukuba, Institute of Biological

Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan

(E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,

Fax:+81-0298-53-6614)

PROJECT = 'Dictyosteleium discoideum cDNA project in Japan'.

Location/Qualifiers





/db\_xref="taxon:44689"

Oy           74 agtactgataacaacaaatggcgcaataaacacaacataa     116  
| | | | |  
Ob       112 acacaccagcgcaggccggaagaacgcccccacaccccccaa     70

/db\_xref="taxon:44689"

```

RESULT 7
C91220      574 bp      mRNA      EST      15-APR-1998
LOCUS       Dictyostellium discoideum slug cDNA, clone SSJ834, mRNA sequence.
DEFINITION
ACCESSION   C91220
NID         93060586
KEYWORDS    EST; EST(expressed sequence tag);
SOURCE      Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
ORGANISM    Dictyostellium discoideum
            Eukaryota; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 574)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostellium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 574)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:gkurushiebank.dna.affrc.go.jp, Tel:++81-0298-53-4664,
            Fax:++81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.
FEATURES    Location/Qualifiers
            source
            1..574
            /organism="Dictyostellium discoideum"
            /strain="AX4"
            /db_xref="taxon:44689"
            /clone_lib="SS"
            /dev_stage="slug"
BASE COUNT  258 a 76 c 63 g 176 t 1 others
ORIGIN

Query Match 5.4%; Score 39.8; DB 28; Length 574;
Best Local Similarity 60.7%; Pred. No. 1.2;
Matches 65; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 38 actctaaacaaacaaagcatcatggaagacagatggagtgacttgataacaaacaatggcg 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 386 AATTGAACAACATATATATATAGAAAGCAACACACACACACACACACACACACACACA 445

QY 98 acaataacaaacaaacaaatgtttacagctcattacagagataattat 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 ACAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAATATAT 492
            (E-mail:gkurushiebank.dna.affrc.go.jp, Tel:++81-0298-53-4664,
            Fax:++81-0298-53-6614)

RESULT 8
A1163084    312 bp      mRNA      EST      03-DEC-1998
LOCUS       A031p65u Hybrid aspen plasmid library Populus tremula x Populus
DEFINITION tremuloides cDNA 5', mRNA sequence.
ACCESSION   A1163084
NID         93854369
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides.
ORGANISM    Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Violales; Salicaceae;
            Populus.
REFERENCE   1 (bases 1 to 312)
AUTHORS     Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A.,
            Holmberg,A., Anini,B., Bhalaria,R., Larsson,M., Villarroel,R., Van
            Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W.,
            Gustafsson,P., Uhlen,M., Sundberg,B. and Lundberg,J.
TITLE       Gene discovery in the wood-forming tissues of poplar: Analysis of
            5,692 expressed sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE     99007314
COMMENT     Contact: Sterky F

```

```

Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se
PCR Primers
FORWARD: AAAGGGGATGTGTCGAAGCG
BACKWARD: GCTTCGGCTCGTATGTGTG
Seq primer: CGTTGTAAACGACGCCAG
High quality sequence stop: 312.
FEATURES    Location/Qualifiers
            1..312
            /organism="Populus tremula x Populus tremuloides"
            /note="Vector: pBluescript SK; Site.1: SalI; Site.2: NotI;
            Cambial region tissues, including developing xylem, the
            meristematic cambial zone and the developing and mature
            phloem, was harvested from 1.5 m actively growing trees.
            cDNA was prepared and cloned into lambda gt2a. DNA was
            isolated and subcloned into pBluescript SK using SalI and
            NotI restriction enzymes."
            /db_xref="taxon:47664"
            /clone_lib="Hybrid aspen plasmid library"
            /tissue_type="cambial region"
            /dev_stage="1.5 m actively growing tree"
            /lab_host="E.coli"
BASE COUNT  113 a 66 c 71 g 62 t
ORIGIN

Query Match 6.4%; Score 39.4; DB 23; Length 312;
Best Local Similarity 65.2%; Pred. No. 1.4;
Matches 58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 187 ccaagatcttatacatgtagctttttaaagggaattagatctctcaagctcttgg 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 CAAAGGGTTTCTCATGCAACTATTGCCAAAGAAATTTATTAGTCTCAACAGGCGTTGGA 259

QY 247 ggacacatgaattcttcataagagatag 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 GGACACCAAAATGCCCATTAAGAGAGAAAG 288

RESULT 9
AU034056    397 bp      mRNA      EST      28-OCT-1998
LOCUS       AU034056 Dictyostellium discoideum slug cDNA, clone SLB863, mRNA sequence.
DEFINITION
ACCESSION   AU034056
NID         93799480
KEYWORDS    EST.
SOURCE      Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM    Dictyostellium discoideum
            Eukaryota; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 397)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostellium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 397)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases.
            Hideko Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:gkurushiebank.dna.affrc.go.jp, Tel:++81-0298-53-4664,
            Fax:++81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.
FEATURES    Location/Qualifiers
            1..397
            /organism="Dictyostellium discoideum"
            /strain="AX4"
            /db_xref="taxon:44689"
            /clone_lib="SL"

```

```

BASE COUNT      207 a      /dev_stage="slug"      69 t      1 others
ORIGIN

Query Match      6.4%; Score 39.4; DB 31; Length 397;
Best Local Similarity 62.9%; Pred. No. 1.5;
Matches 61; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 16 agaaacagtttcaacagtaagctacttcaaaacaaagacatcatcggaagacagatggag 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ACAATCAACTTCACAGCACAATTTTCAACAACAACATTCACACACACCAACACATTTTC 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 76 tacttgtaatacaacaatggcgacaataacaaca 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 AACAAACAACAACAACAACAACAACAACAACAACAACA 219

RESULT 10
N39467      472 bp      mRNA      EST      19-JAN-1996
Yw62d10.r1 Homo sapiens cDNA clone 256819 5'.
N39467
g1162674
EST.
human clone=256819 primer=T7 library=Soares placenta 8to9weeks
2NBHP8to9W vector=PT73D (Pharmacia) with a modified polylinker
host=DH10B (ampicillin resistant) Bst1-Not I Rsite2=Eco RI two
placentae: one from 8 weeks and another from 9 weeks post
conception. 1st strand cDNA was primed with a Not I - oligo(dT)
primer 15'-TGTTACCAATCTGAAGTGGGCGCGCGCGATTTTTTTTTTTT-3',
double-stranded cDNA was size selected, ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I and Eco
RI sites of a modified pMT3 vector (Pharmacia). Library
constructed by Bento Soares and M. Fatima Bonaldo.
Homo sapiens

ORGANISM
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chondata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
1 (bases 1 to 472)
Hillier,D., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 320
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1. .472
/organism="Homo sapiens"
/clone="256819"
<1..>472

BASE COUNT      126 a      94 c      73 g      177 t      2 others
ORIGIN

Query Match      6.3%; Score 39.2; DB 12; Length 472;
Best Local Similarity 52.9%; Pred. No. 1.7;
Matches 83; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```

[illegible]

High quality sequence stops: 337  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
    source  
        1..402  
            /organism="Homo sapiens"  
            /clone="201269"  
            <1..>402  
BASE COUNT     136 a   70 c   75 g   131 t  
ORIGIN

Query Match          6.3%; Score 39; DB 11; Length 402;  
Best Local Similarity 52.8%; Pred.No.1.8;  
Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 265 agaagagatagccattttgagacaatacaccactgagagatatatactgttctctt 324  
     | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 325 AGAAAGAAAAGTACCCCTTGTTCCAAAGACACTGCTCACATTGTTCAGTAATCTTCTG 266  
     | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 325 ctaaaccttaagtgaaccaaaccctaacttttaccttagtcataaaccttagttttca 384  
     | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 265 TTTAGCTGAAGAAAGAACCAATCTGTGTTAATTTTCCATTGAAATAGACTCATAATATGA 206  
     | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 385 agaaaattcccaccttttgaaatgaggaaataggaaaa 423  
     | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 205 AGACAATACTTAACCTTGTTAATCATGCTGGTTGACACA 167  
     | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14  
C22974  
LOCUS C22974 546 bp mRNA EST 27-MAR-1997  
DEFINITION Dictyostelium discoideum gamete cDNA, clone FC-AL20, mRNA sequence.  
ACCESSION C22974  
NID GI922123  
KEYWORDS EST; EST(expressed sequence tag).  
SOURCE Dictyostelium discoideum (strain:KAX3) Sexually mature Gamete cDNA  
to mRNA, clone.lib.FC.  
ORGANISM Dictyostelium discoideum  
Eukaryote; mitochondrial eukaryotes; Dictyosteliida;  
Dictyostelium.  
REFERENCE 1 (bases 1 to 546)  
AUTHORS Urushihara,H.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-1997) to the DDBJ/EMBL/GenBank databases. Hideko  
Urushihara, University of Tsukuba, Institute of Biological  
Sciences; Ten-noudai 1-1-1, Tsukuba-shi, Ibaraki-ken 305, Japan  
(E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,  
Fax:0298-53-6614)  
2 (sites)  
REFERENCE Shimizu,H. and Urushihara,H.  
AUTHORS Sexual-cDNA in CSM  
TITLE Unpublished (1997)  
JOURNAL PROJECT = 'dictyostelium discoideum cDNA project in Japan'.  
COMMENT Location/Qualifiers  
FEATURES  
    source  
        1..546  
            /organism="Dictyostelium discoideum"  
            /strain="KAX3"  
            /db\_xref="taxon:44689"  
            /cell\_type="Gamete"  
            /clone\_lib="FC"  
            /dev\_stage="Sexually mature"  
BASE COUNT 317 a 110 c 33 g 85 t 1 others  
ORIGIN

Query Match          6.3%; Score 39; DB 17; Length 546;  
Best Local Similarity 59.5%; Pred.No.2;  
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 6 aggcatggatagaacagtttcaacagtaagtcttcaaaacaaagcatcgccaag 65  
     | | | | | | | | | | | | | | | | | | | | | | | | | |



THIS PAGE BLANK (USPTO)

\*\*\*\*\*

WQESRL (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jul 14 17:26:36 1999; MasPar time 15.20 Seconds  
Tabular output not generated. 288.149 Million cell updates/sec

Title: >US-09-156-580-2  
Description: (1-206) from US09156580A.pap  
Perfect Score: 1495  
Sequence: 1 MAGMDRNSFNSYFNKNSIM.....IGLISESKEDLDLELRGSG 206

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 31.942; Variance 133.844; scale 0.239

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	119	8.0	227	20	Arabidopsis Strz polyp	1.81e-01
2	115	7.7	2391	10	Carbamoyl-phosphate-s	3.78e-01
3	114	7.6	570	26	Human Fchd531 gene pr	4.54e-01
4	111	7.4	31	37	Fragment of human sec	7.85e-01
5	111	7.4	183	37	Fragment of human sec	7.85e-01
6	111	7.4	203	37	Human secreted protei	7.85e-01
7	101	6.8	364	20	E. coli O157:H7 F12 r	4.70e+01
8	97	6.5	88	38	A nucleic acid bindin	9.48e+00
9	97	6.5	88	38	An anti-HIV zinc fing	9.48e+00
10	94	6.3	1060	30	Homo sapiens 20q13 am	1.59e+01
11	92	6.2	423	17	Transcription factor-	2.24e+01
12	92	6.2	706	12	BCL-6 zinc finger pro	2.24e+01
13	93	6.2	803	37	Myc-binding zinc-fing	1.89e+01
14	91	6.1	547	17	HIC-1 polypeptide.	2.66e+01
15	91	6.1	957	22	H. pylori transmembra	2.66e+01
16	91	6.1	2496	26	Mutant YLR087c protei	2.66e+01

17	91	6.1	2942	25	W22050	Saccharomyces cerevis	2.66e+01
18	89	6.0	92	18	R92699	Zinc finger protein Z	3.73e+01
19	89	6.0	1719	16	R92100	Human RIZ allele D283	3.73e+01
20	88	5.9	343	16	R77656	Human zinc finger pro	4.42e+01
21	88	5.9	566	37	W82559	A. thaliana homologou	4.42e+01
22	88	5.9	675	17	R94903	Human lastin.	4.42e+01
23	88	5.9	1706	16	R92103	Rat RIZ.	4.42e+01
24	87	5.8	59	12	R63135	Egr-1 nuclear localiz.	5.22e+01
25	87	5.8	89	11	R63136	Egr-1 polynucleotide	5.22e+01
26	87	5.8	91	39	W84456	Zif268 three finger p	5.22e+01
27	87	5.8	91	15	R83492	Zif268 zinc finger do	5.22e+01
28	87	5.8	153	39	W84391	Zif268-Jun construct.	5.22e+01
29	87	5.8	153	39	W84392	Zif268-Fos construct.	5.22e+01
30	87	5.8	153	15	R83488	Zif268-Fos zinc finge	5.22e+01
31	87	5.8	153	15	R83487	Zif268-Jun zinc finge	5.22e+01
32	87	5.8	181	39	W84394	The zif268-zif268 con	5.22e+01
33	87	5.8	181	15	R83490	Zif268-zif268 zinc fi	5.22e+01
34	86	5.8	387	37	W76987	Human EGR-3 protein.	6.17e+01
35	87	5.8	455	37	W76986	Human EGR-2 protein.	5.22e+01
36	87	5.8	456	1	P93114	EGR2.	5.22e+01
37	87	5.8	456	11	R63130	Human Egr-2.	5.22e+01
38	87	5.8	496	37	W76984	Mouse Egr-1 protein.	5.22e+01
39	87	5.8	500	32	W37948	Amino acid sequence o	5.22e+01
40	86	5.8	507	22	W15574	Mouse Aiolos polypept	6.17e+01
41	87	5.8	533	11	R63129	Mouse Egr-1 clone OC3	5.22e+01
42	87	5.8	533	1	P93113	Egr-1.	5.22e+01
43	87	5.8	543	37	W76985	Human EGR-1 protein.	5.22e+01
44	87	5.8	543	1	R24022	Human promyelo-leukae	5.22e+01
45	86	5.8	791	17	R95566	N. meningitidis serot	6.17e+01

## ALIGNMENTS

RESULT 1  
ID W01604 standard; Protein; 227 AA.  
AC W01604;  
DE 16-APR-1997 (first entry)  
DE Arabidopsis Strz polypeptide.  
KW Strz; STO; salt tolerance; transgenic plant; yeast.  
OS Arabidopsis thaliana.  
PN W09639020-A1.  
PD 12-DEC-1996.  
PI 31-MAY-1996; U08095.  
PR 08-JUN-1995; US-471717.  
PA (REGC ) UNIV CALIFORNIA.  
PI Gasser CS, Lippuner V;  
DR WPI: 97-042725/04.  
DR N-PSDB: T58281  
PT Nucleic acids encoding Strz and STO polypeptide(s) which confer salt  
PT tolerance - useful for producing transgenic plants or other  
PT organisms, e.g. yeast, able to grow under saline conditions  
PS Claim 2; Page 22; 31pp; English.  
CC The Strz (W01604) and STO (W01605) polypeptides of Arabidopsis  
CC thaliana have the ability to confer tolerance to saline conditions  
CC on plants, yeast, fungi and other organisms. They can be obtd. by  
CC expression of Strz (T58281) and STO (T58282) cDNA clones that were  
CC isolated from an A. thaliana library by complementation of a salt  
CC tolerance phenotype in yeast. Recombinant expression cassettes  
CC can be used to express the Strz and STO polypeptides in transgenic  
CC plants, allowing the plants to grow under saline conditions.  
SQ Sequence 227 AA;

Query Match 8.0%; Score 119; DB 20; Length 227;  
Best Local Similarity 55.6%; Pred. No. 1.81e-01;  
Matches 15; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 79 syksvcdkffssyqalggghkashrkn 105

QY 65 SYTSCFKRFRSAQALGGHMHVRRD 91

RESULT 2  
ID R55694 standard; Protein; 2391 AA.

AC R55694;  
 DT 06-DEC-1994 (first entry)  
 DE Carbamoyl-phosphate-synthetase II.  
 KW Carbamoyl-phosphate-synthetase II; CPSII; pscpsii gene;  
 KW malaria.  
 OS Plasmodium falciparum.  
 FH Key Location/Qualifiers  
 FT domain 1..690  
 FT domain /note= "glutamine-amidotransferase domain"  
 FT domain 1..270  
 FT domain /note= "structural subdomain"  
 FT peptide 271..482  
 FT domain /note= "insert sequence"  
 FT domain 483..690  
 FT domain /note= "glutaminase subdomain"  
 FT domain 691..2391  
 FT domain /note= "carbamoyl-phosphate-synthase domain"  
 FT domain 691..1254  
 FT domain /note= "ATP binding subdomain CPSa"  
 FT peptide 1255..1857  
 FT domain /note= "insert sequence"  
 FT domain 1858..2391  
 FT domain /note= "ATP binding subdomain CPSb"  
 PN WO9412643-A.  
 PD 09-JUN-1994.  
 PF 02-DEC-1993; AU0617.  
 PR 03-DEC-1992; AU-006206.  
 PR 16-DEC-1992; AU-006380.  
 PI (UNIX ) UNISEARCH LTD.  
 PA Flores MV, Osullivan WJ, Stewart TS;  
 DR WPI: 94-200271/24.  
 DR N-PSDB; Q62924.  
 PT Nucleic acid encoding carbamoyl phosphate synthetase II -  
 PT isolated from Plasmodium falciparum, used to develop prods. for  
 PT the treatment of malaria.  
 PS Disclosure: Page 6-16; 31pp; English.  
 CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II  
 CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes  
 CC a protein that includes 2 insert sequences not found in other CPSII  
 CC proteins. The first separates the putative structural subdomain and  
 CC the glutaminase subdomain of the glutamine-amidotransferase subunit  
 CC of CPSII, while the second separates 2 ATP binding subdomains of the  
 CC CPSII subunit, CPSa and CPSb.  
 SQ Sequence 2391 AA;  
 Query Match 7.7%; Score 115; DB 10; Length 2391;  
 Best Local Similarity 47.6%; Pred. No. 3,78e-01;  
 Matches 20; Conservative 10; Mismatches 10; Indels 2; Gaps 2;  
 Db 1159 leqlsfndlkfkkhgfsgkqahyisfntsdnnnnnnmiss 1200  
 QY :::: |||| :::: ||:: ||:: ||||| ||  
 4 MDRNSFNS-KYFNKKSIMARQME-YLNNGNGDNNNNNNVTSS 43  
 RESULT 3  
 ID W36002 standard; Protein: 570 AA.  
 AC W36002;  
 DE Human Fchd531 (first entry)  
 KW Fchd531 gene; differential expression; endothelial cell; human;  
 KW shear stress; cardiovascular disease; atherosclerosis; ischaemia;  
 KW reperfusion; hypertension; restenosis; arterial inflammation;  
 KW therapy; diagnosis; drug screening; marker.  
 OS Homo sapiens.  
 PN WO9730065-A1.  
 PD 21-AUG-1997.  
 PF 14-FEB-1997; U02291.  
 PR 13-FEB-1997; US-799910.  
 PR 16-FEB-1996; US-011787.  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PI Falb DA;  
 DR WPI: 97-424966/39.  
 DR N-PSDB; T94467.

PT New genes differentially expressed in cardiovascular disease - used  
 PT for diagnosis, drug screening and treatment of cardiovascular  
 PT disease, e.g. atherosclerosis, restenosis, hypertension, etc  
 PS Example 7; Fig 1A-1B; 163pp; English.  
 CC This protein is encoded by the novel human fchd531 gene (see  
 CC T94467) that is down-regulated in endothelial cells subjected to  
 CC turbulent and laminar shear stress. Shear stress is thought to be  
 CC responsible for the prevalence of atherosclerotic lesions in areas  
 CC of unusual circulatory flow. The fchd531 gene product has 94%  
 CC similarity to the mouse penta zinc finger gene (Pzf). Gene  
 CC products were also identified for novel genes fchd540 (see  
 CC W36003) and fchd545 (see W36004), which are respectively up- and  
 CC down-regulated in endothelial cells subjected to shear stress.  
 CC Novel fchd531, fchd540, fchd602 and fchd605 genes (see  
 CC T94467-71) provide a fingerprint for the study of cardiovascular  
 CC diseases, including atherosclerosis, ischaemia/reperfusion,  
 CC hypertension, restenosis and arterial inflammation. Methods are  
 CC provided for the diagnosis, monitoring in clinical trials,  
 CC screening for therapeutically effective compounds, and treatment of  
 CC cardiovascular diseases based on discoveries regarding the  
 CC expression patterns of these novel genes.  
 SQ Sequence 570 AA;  
 Query Match 7.6%; Score 114; DB 26; Length 570;  
 Best Local Similarity 32.3%; Pred. No. 4,54e-01;  
 Matches 10; Conservative 10; Mismatches 11; Indels 0; Gaps 0;  
 Db 370 rdyiceycarafkshnlavhrmhtgkpl 400  
 QY ||||| ||||| :::: ||:: ||:: ||||| ||  
 64 RSYTCSFKRFRSAQAAGGHMNVHRDRAI 94  
 RESULT 4  
 ID W75038 standard; Protein: 31 AA.  
 AC W75038;  
 DT 25-JAN-1999 (first entry)  
 DE Fragment of human secreted protein encoded by gene 64.  
 KW Human; secreted protein; testis; tumour; foetal brain tissue;  
 KW fusion protein; cancer; central nervous system; seizure;  
 KW diagnosis; neurodegenerative disease.  
 OS Homo sapiens.  
 PN WO9839448-A2.  
 PD 11-SEP-1998.  
 PF 06-MAR-1998; U04493.  
 PR 02-OCT-1997; US-061060.  
 PR 07-MAR-1997; US-038621.  
 PR 07-MAR-1997; US-040161.  
 PR 07-MAR-1997; US-040162.  
 PR 07-MAR-1997; US-040163.  
 PR 07-MAR-1997; US-040333.  
 PR 07-MAR-1997; US-040334.  
 PR 07-MAR-1997; US-040336.  
 PR 07-MAR-1997; US-040626.  
 PR 11-APR-1997; US-043311.  
 PR 11-APR-1997; US-043312.  
 PR 11-APR-1997; US-043313.  
 PR 11-APR-1997; US-043314.  
 PR 11-APR-1997; US-043568.  
 PR 11-APR-1997; US-043569.  
 PR 11-APR-1997; US-043576.  
 PR 11-APR-1997; US-043578.  
 PR 11-APR-1997; US-043580.  
 PR 11-APR-1997; US-043669.  
 PR 11-APR-1997; US-043670.  
 PR 11-APR-1997; US-043671.  
 PR 11-APR-1997; US-043672.  
 PR 11-APR-1997; US-043674.  
 PR 23-MAY-1997; US-047492.  
 PR 23-MAY-1997; US-047500.  
 PR 23-MAY-1997; US-047501.  
 PR 23-MAY-1997; US-047502.  
 PR 23-MAY-1997; US-047503.  
 PR 23-MAY-1997; US-047581.



PR	23-MAY-1997	US-047580
PR	23-MAY-1997	US-047581
PR	23-MAY-1997	US-047582
PR	23-MAY-1997	US-047583
PR	23-MAY-1997	US-047584
PR	23-MAY-1997	US-047585
PR	23-MAY-1997	US-047586
PR	23-MAY-1997	US-047587
PR	23-MAY-1997	US-047588
PR	23-MAY-1997	US-047589
PR	23-MAY-1997	US-047590
PR	23-MAY-1997	US-047592
PR	23-MAY-1997	US-047593
PR	23-MAY-1997	US-047594
PR	23-MAY-1997	US-047595
PR	23-MAY-1997	US-047596
PR	23-MAY-1997	US-047597
PR	23-MAY-1997	US-047598
PR	23-MAY-1997	US-047599
PR	23-MAY-1997	US-047600
PR	23-MAY-1997	US-047601
PR	23-MAY-1997	US-047612
PR	23-MAY-1997	US-047613
PR	23-MAY-1997	US-047614
PR	23-MAY-1997	US-047615
PR	23-MAY-1997	US-047617
PR	23-MAY-1997	US-047618
PR	23-MAY-1997	US-047632
PR	23-MAY-1997	US-047633
PR	06-JUN-1997	US-048964
PR	06-JUN-1997	US-048974
PR	13-JUN-1997	US-049610
PR	08-JUL-1997	US-051926
PR	16-JUL-1997	US-052874
PR	18-AUG-1997	US-055724
PR	22-AUG-1997	US-056630
PR	22-AUG-1997	US-056631
PR	22-AUG-1997	US-056632
PR	22-AUG-1997	US-056633
PR	22-AUG-1997	US-056637
PR	22-AUG-1997	US-056662
PR	22-AUG-1997	US-056664
PR	22-AUG-1997	US-056845
PR	22-AUG-1997	US-056862
PR	22-AUG-1997	US-056864
PR	22-AUG-1997	US-056872
PR	22-AUG-1997	US-056874
PR	22-AUG-1997	US-056875
PR	22-AUG-1997	US-056876
PR	22-AUG-1997	US-056877
PR	22-AUG-1997	US-056878
PR	22-AUG-1997	US-056879
PR	22-AUG-1997	US-056880
PR	22-AUG-1997	US-056881
PR	22-AUG-1997	US-056882
PR	22-AUG-1997	US-056884
PR	22-AUG-1997	US-056886
PR	22-AUG-1997	US-056887
PR	22-AUG-1997	US-056888
PR	22-AUG-1997	US-056889
PR	22-AUG-1997	US-056903
PR	22-AUG-1997	US-056909
PR	22-AUG-1997	US-056910
PR	22-AUG-1997	US-056911
PR	05-SEP-1997	US-057650
PR	05-SEP-1997	US-057659
PR	12-SEP-1997	US-057761
PR	12-SEP-1997	US-057875
PA	(HUMA-) HUMAN GENOME SC	
PI	Bednarik DP, Brewer LA, Pi	
PI	Feng P, Ferrie AM, Fisc	

Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,  
Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
WPI; 98-506364/A3.  
New isolated human genes and the secreted polypeptide(s) they encode  
- useful for diagnosis and treatment of e.g. cancers, neurological  
disorders, immune diseases, inflammation or blood disorders  
This disclosure; Page 43; 721pp; English.  
This sequence represents a fragment of a secreted human protein encoded  
by the nucleic acid molecule designated Gene 64 (V59574).  
The gene can be used to generate fusion proteins by linking to the gene  
stability of the fused protein as compared to the human protein only.  
The invention relates to 186 novel genes and their fragments (nucleic  
acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which  
are useful for preventing, treating or ameliorating medical conditions  
e.g. by protein or gene therapy. Also, pathological conditions can be  
diagnosed by determining the amount of the new polypeptides in a sample  
or by determining the presence of mutations in the new polynucleotides.  
Specific uses are described for each of the 186 polynucleotides, based on  
which tissues they are most highly expressed in (see v59511 for described  
uses).

Query Match                      7.4%; Score 111; DB 37; Length 31:  
Best Local Similarity 34.5%; Pred. No. 7.85e-01;  
Matches 10; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

Dbb 3 rdyiceycarafksshnlavrhmtgek 31  
| | | | | : | | | | | : | | : | :  
Qy 64 RSYTSCFKREFRSQAQLGGHNMVRRDR 92

RESULT 5

ID W75040 standard; Protein; 183 AA.

AC W75040;

DE 25-JAN-1999 (first entry)

DT Fragment of human secreted protein encoded by gene 64.

KW Human; secreted protein; testis; tumour; foetal brain tissue;

KW fusion protein; cancer; central nervous system; seizure;

KW diagnosis; neurodegenerative disease.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 58 /label= unknown

FT FT 78 /label= unknown

FT Misc-difference 98 /label= unknown

FT FT 150 /label= unknown

FT FT 183 /label= unknown

FT FT /label= unknown

PD WO9839448-A2.

PN 11-SEP-1998.

PP 06-MAR-1998; UO4493.

PP 02-OCT-1997; US-061060.

PR 07-MAR-1997; US-038621.

PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.

PR 07-MAR-1997; US-040163.

PR 07-MAR-1997; US-040333.

PR 07-MAR-1997; US-040334.

PR 07-MAR-1997; US-040336.

PR 07-MAR-1997; US-040826.

PR 11-APR-1997; US-043311.

PR 11-APR-1997; US-043312.

PR 11-APR-1997; US-043313.

PR 11-APR-1997; US-043314.

PR 11-APR-1997; US-043568.

PR 11-APR-1997; US-043569.

PR 11-APR-1997; US-043576.

PR 11-APR-1997; US-043578.

PR 11-APR-1997; US-043580.

PR	11-APR-1997;	US-043669.
PR	11-APR-1997;	US-043670.
PR	11-APR-1997;	US-043671.
PR	11-APR-1997;	US-043672.
PR	11-APR-1997;	US-043673.
PR	23-MAY-1997;	US-047482.
PR	23-MAY-1997;	US-047500.
PR	23-MAY-1997;	US-047501.
PR	23-MAY-1997;	US-047502.
PR	23-MAY-1997;	US-047503.
PR	23-MAY-1997;	US-047581.
PR	23-MAY-1997;	US-047582.
PR	23-MAY-1997;	US-047583.
PR	23-MAY-1997;	US-047584.
PR	23-MAY-1997;	US-047585.
PR	23-MAY-1997;	US-047586.
PR	23-MAY-1997;	US-047587.
PR	23-MAY-1997;	US-047588.
PR	23-MAY-1997;	US-047589.
PR	23-MAY-1997;	US-047590.
PR	23-MAY-1997;	US-047592.
PR	23-MAY-1997;	US-047593.
PR	23-MAY-1997;	US-047594.
PR	23-MAY-1997;	US-047595.
PR	23-MAY-1997;	US-047596.
PR	23-MAY-1997;	US-047597.
PR	23-MAY-1997;	US-047598.
PR	23-MAY-1997;	US-047599.
PR	23-MAY-1997;	US-047600.
PR	23-MAY-1997;	US-047601.
PR	23-MAY-1997;	US-047612.
PR	23-MAY-1997;	US-047613.
PR	23-MAY-1997;	US-047614.
PR	23-MAY-1997;	US-047615.
PR	23-MAY-1997;	US-047617.
PR	23-MAY-1997;	US-047618.
PR	23-MAY-1997;	US-047632.
PR	23-MAY-1997;	US-047633.
PR	06-JUN-1997;	US-048964.
PR	06-JUN-1997;	US-048974.
PR	13-JUN-1997;	US-049610.
PR	08-JUL-1997;	US-051926.
PR	16-JUL-1997;	US-052874.
PR	18-AUG-1997;	US-055724.
PR	22-AUG-1997;	US-056630.
PR	22-AUG-1997;	US-056631.
PR	22-AUG-1997;	US-056632.
PR	22-AUG-1997;	US-056636.
PR	22-AUG-1997;	US-056637.
PR	22-AUG-1997;	US-056662.
PR	22-AUG-1997;	US-056664.
PR	22-AUG-1997;	US-056845.
PR	22-AUG-1997;	US-056862.
PR	22-AUG-1997;	US-056864.
PR	22-AUG-1997;	US-056872.
PR	22-AUG-1997;	US-056873.
PR	22-AUG-1997;	US-056875.
PR	22-AUG-1997;	US-056876.
PR	22-AUG-1997;	US-056877.
PR	22-AUG-1997;	US-056878.
PR	22-AUG-1997;	US-056879.
PR	22-AUG-1997;	US-056880.
PR	22-AUG-1997;	US-056881.
PR	22-AUG-1997;	US-056882.
PR	22-AUG-1997;	US-056884.
PR	22-AUG-1997;	US-056886.
PR	22-AUG-1997;	US-056887.
PR	22-AUG-1997;	US-056888.
PR	22-AUG-1997;	US-056889.
PR	22-AUG-1997;	US-056892.
PR	22-AUG-1997;	US-056893.
PR	22-AUG-1997;	US-056894.
PR	22-AUG-1997;	US-056903.
PR	22-AUG-1997;	US-056904.
PR	22-AUG-1997;	US-056905.
PR	22-AUG-1997;	US-056906.
PR	22-AUG-1997;	US-056907.
PR	22-AUG-1997;	US-056908.
PR	22-AUG-1997;	US-056909.
PR	22-AUG-1997;	US-056910.
PR	22-AUG-1997;	US-056911.
PR	05-SEP-1997;	US-057650.
PR	05-SEP-1997;	US-057659.
PR	05-SEP-1997;	US-057761.
PR	12-SEP-1997;	US-058785.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,	
PI	Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,	
PI	Kyav H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,	
PI	Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;	
DR	WPI; 98-506364/43.	
PT	New isolated human genes and the secreted polypeptide(s) they encode	
PT	- useful for diagnosis and treatment of e.g. cancers, neurological	
PT	disorders, immune diseases, inflammation or blood disorders	
PS	Disclosure; Page 43; 72lpp; English.	
CC	This sequence represents a fragment of a secreted human protein encoded	
CC	by the nucleic acid molecule designated gene 64 (V59574).	
CC	The gene can be used to generate fusion proteins by linking to the gene	
CC	to a human immunoglobulin FC portion (e.g. V59502) for increasing the	
CC	stability of the fused protein as compared to the human protein only.	
CC	The invention relates to 186 novel genes and their fragments (nucleic	
CC	acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which	
CC	are useful for preventing, treating or ameliorating medical conditions	
CC	e.g. by protein or gene therapy. Also, pathological conditions can be	
CC	diagnosed by determining the amount of the new polypeptides in a sample	
CC	or by determining the presence of mutations in the new polynucleotides.	
CC	Specific uses are described for each of the 186 polynucleotides, based on	
CC	which tissues they are most highly expressed in (see V59511 for described	
CC	uses).	
SQ	Sequence 183 AA;	
Query Match	7.4%; Score 111; DB 37; Length 183;	
Best Local Similarity	34.5%; Pred. No. 7.85e-01;	

PR	22-AUG-1997	US-056908.
PR	22-AUG-1997	US-056909.
PR	22-AUG-1997	US-056910.
PR	22-AUG-1997	US-056911.
PR	05-SEP-1997	US-057650.
PR	05-SEP-1997	US-057659.
PR	05-SEP-1997	US-057761.
PR	12-SEP-1997	US-058785.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,	
PI	Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,	
PI	Kyav H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,	
PI	Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;	
DR	WPI: 98-506364/43.	
PT	New isolated human genes and the secreted polypeptide(s) they encode	
PT	- useful for diagnosis and treatment of e.g. cancers, neurological	
PT	disorders, immune diseases, inflammation or blood disorders	
PS	Disclosure; Page 43; 721pp; English.	
CC	This sequence represents a fragment of a secreted human protein encoded	
CC	by the nucleic acid molecule designated gene 64 (V59574).	
CC	The gene can be used to generate fusion proteins by linking to the gene	
CC	to a human immunoglobulin FC portion (e.g. V59502) for increasing the	
CC	stability of the fused protein as compared to the human protein only.	
CC	The invention relates to 186 novel genes and their fragments (nucleic	
CC	acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which	
CC	are useful for preventing, treating or ameliorating medical conditions	
CC	e.g. by protein or gene therapy. Also, pathological conditions can be	
CC	diagnosed by determining the amount of the new polypeptides in a sample	
CC	or by determining the presence of mutations in the new polynucleotides.	
CC	Specific uses are described for each of the 186 polynucleotides, based on	
CC	which tissues they are most highly expressed in (see V59511 for described on	
CC	uses).	
SQ	Sequence	183 AA;
Query Match 7.4%; Score 111; DB 37; Length 183;		
Best Local Similarity 34.5%; Pred. No. 7.85e-01;		
Matches 10; Conservative 8; Mismatches 11; Indels 0; Gaps 0;		
Db	152 rdviceycarafskshnlavhrmhtgek 180	
Qy	64 RSYTCSEKREFRSAQALGGHNNVRRDR 92	
RESULT 6		
ID	W74793 standard; Protein; 203 AA.	
AC	W74793;	
DT	25-JAN-1999 (first entry)	
DE	Human secreted protein encoded by gene 64 clone HMWEX24.	
KW	Human; secreted protein; testis; tumour; foetal brain tissue;	
KW	fusion protein; cancer; central nervous system; seizure;	
KW	diagnosis; neurodegenerative disease.	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	Misc-difference 203	label= unknown
FT	W09839448-A2.	
PD	11-SEP-1998.	
PF	06-MAR-1998; U04493.	
PR	02-OCT-1997; US-061060.	
PR	07-MAR-1997; US-038621.	
PR	07-MAR-1997; US-040161.	
PR	07-MAR-1997; US-040162.	
PR	07-MAR-1997; US-040163.	
PR	07-MAR-1997; US-040333.	
PR	07-MAR-1997; US-040334.	
PR	07-MAR-1997; US-040336.	
PR	07-MAR-1997; US-040626.	
PR	11-APR-1997; US-043311.	
PR	11-APR-1997; US-043312.	
PR	11-APR-1997; US-043313.	
PR	11-APR-1997; US-043314.	
PR	11-APR-1997; US-043568.	
PR	11-APR-1997; US-043569.	



CC product) of Vibrio cholerae. They are considered candidate  
CC immunogens for preparing antibody reagents for detection of the  
CC O157:H7 strain.  
SQ Sequence 364 AA;

```

Query Match      6.8%  Score 101:  DB 20:  Length 364;
Best Local Similarity 23.3%:  Pred. No. 4.70e+00;
Matches 24:  Conservative 17:  Mismatches 35:  Indels 6:  Gaps 6:

Db 283 sllrtaeaeelrnlhldkllietrvfyvptmpmysekyqkhtaediwigilpsf 342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 93 AIL-RQSPDRINRYSLLNLNLENPNFYPSHN-PSFSRKEPPEMRK-LG-KGV-VPN- 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 343 pslnsqevlyicesinefyysdk 364
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 NHLKRSARGRFGEKIDSFMOEK 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8  
ID W87557 standard: Proteinp: 88 AA

AC	W87557;
AD	01-MAR-1999 (first entry)
DE	A nucleic acid binding protein.
DE	Nucleic acid binding protein; Cys2-His2 zinc finger class; diagnosis;
KW	genetic disorder; chimeric restriction enzyme; gene therapy.
KW	Synthetic.
OS	WO9853058-A1.
PN	WO9853058-A1.
PD	26-NOV-1998.
PD	26-MAY-1998; G01512.
PF	23-MAY-1997; GB-010809.
PR	(MEDI-) MEDICAL RES COUNCIL.
PA	Choo Y, Isalan M, Klug A;
PI	WPI; 99-024578/02.
DR	N-PSDB: V83634.
DR	PT-Preparing Cys2-His2 zinc finger class nucleic acid binding proteins
PT	PT - capable of binding to a nucleic acid quadruplet, by mutating sites
PT	PT in model zinc finger domains according to a defined set of
PT	PT substitutions

The present sequence represents a nucleic acid binding protein (NBP) of the Cys2-His2 zinc finger class. The NBP is capable of binding to a nucleic acid quadruplet in a target sequence. The specification describes a method of preparing the NBP. The method is used for designing nucleic acid binding proteins which are useful in medicine. The proteins are specifically engineered to recognise particular nucleic acid sequences and as such are suitable for diagnosis of genetic disorders. The proteins can be used in the manufacture of chimeric restriction enzymes, in which a nucleic acid cleaving domain is fused to a nucleic acid binding domain comprising a zinc finger. Fusion proteins comprising NBP and an integrase, e.g. viral integrase, can be used to target nucleic acid sequences in vivo. In gene therapy applications, the method may be targeted to the delivery of functional genes into defective genes, or the delivery of nonsense nucleic acid in order to disrupt undesired nucleic acid. Genes may also be delivered to known, repetitive stretches of nucleic acid, e.g. centromeres, together with an activating sequence such as an LCR. NBP can be specifically used to knockout cells having mutant proteins e.g. mutant ras. They can also be used to modulate the action of transcription factors, e.g. the activity of HIV tat may be reduced by NBP specific for HIV TAR. NBP may also be coupled to toxic molecules, e.g. nucleases, which are capable of selectively destroying cells which comprise a mutation in their endogenous nucleic acid.

Query Match	6.5%	Score 97;	DB 38;	Length 88;
Best Local Similarity	28.0%	Pred. No. 9.48e+00;		
Matches	14;	Conservative 16;	Mismatches 18;	Indels 2;
				Gaps 2;

Db 40 mr-nfrs<sup>gn</sup>l<sup>tr</sup>hl<sup>tr</sup>htgkpfq<sup>cr</sup>icm<sup>rn</sup>frqadhl<sup>q</sup>ehl<sup>k</sup>htgk 88  
: | : : | : | : : : | : | : | : | : | : :  
Qv 44 LRDYGNEDHLGLG<sup>LF</sup>SWPP-RSYTCSEKREFRSAQALGGHN<sup>VH</sup>RRDR 92

**RESULT** 9

ID WB4299 standard; Protein; 88 AA.  
AC WB4299;  
DT 18-MAR-1999 (first entry)  
DE An anti-HIV zinc finger.  
KW Anti-HIV zinc finger; nucleic acid  
KW Cys2-His2 zinc finger; detection;  
OS Synthetic.  
PN W09853059-A1.  
PP 26-NOV-1998.  
PF 26-MAY-1998; G01514  
PR 23-MAY-1997; GB-010807.  
PR (MEDI-) MEDICAL RES COUNCIL,  
PI Choo Y, Isalan M, Klug A;  
PI WPI: 99-045308/04.  
DR N-PSDB: V99467.  
PT Preparation of nucleic acid binding  
PT sequences of a Cys2-His2 zinc finger  
PT base triplet in a target nucleic acid  
PS Example 5; Page 42; 62pp; English.  
CC The present sequence encodes an anti-  
CC was made to exemplify the invention  
CC method for preparing a nucleic acid  
CC Cys2-His2 zinc finger class capable  
CC triplet in a target nucleic acid  
CC the triplet by an alpha-helical  
CC in the protein is determined as fo  
CC triplet is A, then position +6 in  
CC Val; (b) if the 5' base in the tri  
CC alpha-helix is Ser, Thr, Val, Ala  
CC for designing a protein which is c  
CC nucleic acid sequence. The NABPs  
CC nucleic acid molecules. They can a  
CC the delivery of functional genes i  
CC of nonsense nucleic acid to disrupt  
SO Sequence 88 AA.

Query Match 6.5%; Score 97; DB 39; Length 88;  
Best Local Similarity 28.0%; Pred. No. 9.48e+00;  
Matches 14; Conservative 16; Mismatches 18; Indels 2; Gaps 2

Db	40 mr-nfsrsdnlrhlrlthtgekpfqcrimrnfrqadhlqlchlkhtgk 88 :   ::   ::   : ::   :   :   :   :   :   :   :   :   :
Qv	44 LRDNYGNEHDILGGLEFWPP-RSYTCSECKREERSAALGGHNNVHRDR 92

RESULT	10
ID	W3975 standard; Protein; 1060 AA.
AC	Z3975;
DT	21-JUL-1998 (first entry)
DE	Homo sapiens 20q13 amplicon ZABC-L
KW	20q13 amplicon; chromosome 20; tumor
KW	chromosomal abnormalities; probe;
KW	treatment; age-related macular deg
KW	Leber's congenital amaurosis; zinc
OS	Homo sapiens.
FH	Key Location/Qualifier
FT	Misc_difference 163 /note= "undefined"
FT	Misc_difference 164 /note= "undefined"
FT	Misc_difference 165 /note= "undefined"
FT	Misc_difference 166 /note= "undefined"
PN	WO9802539-A1.

PR 13-JUL-1997; US-785343.  
 PR 17-JAN-1997; US-785532.  
 PR 15-JUL-1996; US-680395.  
 PR 16-OCT-1996; US-731499.  
 PA (REGC) UNIV CALIFORNIA.  
 PI Collins CC, Godfrey T, Gray JW, Hwang SI, Kowbel D.  
 PI Rommens J.

```

Query Match          6.28; Score 92; DB 12; Length 706;
Best Local Similarity 30.08; Pred. No. 2.24e+01;
Matches 9; Conservative 10; Mismatches 11; Indels 0; Caps

```

---

```

Db      628 kypyceicgtrfrhqltkshlrhtgkcp 657
      : : : : : : : : : : : : : : : : : : : :
QY      64 RSYTSCFKREFRSAQALGGHNVHRDRA 93

```

---

```

RESULT 13
ID      W81756 standard; Protein; 803 AA.
AC      W81756;
DT      27-JAN-1999 (first entry)
DE      Myc-binding zinc-finger protein.
KW      Myc; zinc-finger protein; transcription modulation; gene therapy.
OS      Unknown.
PN      EP-875567-A2.

```

ID	W20841 standard; Protein; 957 AA.
AC	W20841;
DT	17-JUL-1997 (first entry)
DE	H. pylori transmembrane protein, 12ael0622orf16.
HE	Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW	identification; binding compound; bacterium; life cycle; activator;
KW	bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW	diagnosis.
OS	Helicobacter pylori.
PN	W09640893-AL.
PD	19-DEC-1996.
PF	06-JUN-1996; U091122.
PR	07-JUN-1995; US-487032.
PR	01-APR-1996; US-630405.
PA	(ASTR ) ASTRA AB.
PI	Berglindh OT, Smith D, Mellgaard BL;
DR	WPI; 97-052306/05.
DR	N-PSDB: T68094.
PT	Helicobacter pylori nucleic acid sequences and related
PT	polypeptides(s) - useful for vaccines to treat or prevent H. pylori
PT	infection, and to detect Helicobacter
PS	Claim 73; Page 1249-51; 1481pp; English.
CC	This sequence represents a H. pylori protein likely to contain two
CC	membrane spanning regions.
CC	The protein may be used in a vaccine to prevent or treat H. pylori
CC	infection or to identify H. pylori polypeptide binding compounds,
CC	useful as potential H. pylori life cycle activators or inhibitors.
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from
CC	overlapping contigs generated by mechanically shearing the bacterial
CC	DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC	and the predicted coding regions defined by computer evaluation. To
CC	identify likely H. pylori antigens for vaccine development, the amino
CC	acid sequences predicted from various ORF were analysed for significant
CC	homology to other known or exported membrane proteins. Having identified
CC	and determined the sequences of interest, particular regions can be
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide
CC	production, e.g. in E. coli hosts.
SQ	Sequence 957 AA;

CC This sequence represents a H. pylori protein likely to contain two  
CC membrane spanning regions.  
CC  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 957 AA;

```
CC    and the predicted coding regions determined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence      957 AA;  
  
Query Match          6.1%; Score 91; DB 22; Length 957;  
Best Local Similarity 24.7%; Pred. No. 2.86E+01;  
Matches     20; Conservative   22; Mismatches   36; Indels   3; Gaps   3;  
  
Db       74 eskkefyafqkghystancdafakklnhlksfhkcvenhfenfinrlfaymatgsgk 133  
QY        |::|::|::|::|::|::|::|::|::|::|::|::|::|:  
114 EPNPNFYPSINPFSRKFPPFERMKLGKGVPVNNHLSKARGRF-G-VEKDSEFMQEKECT 171  
  
Db       134 tiiviikvellsvamgmgiipr 154  
QY        |||||::|::|::|::|  
172 TTVIKK-SPEFLRLDLGIGLIS 191  
  
Search completed: Wed Jul 14 17:26:55 1999  
Job time : 19 secs.
```

RESULT	15
--------	----

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```
Run on:      Wed Jul 14 17:26:04 1999;  MasPar time 11.70 Seconds
           705.211 Million cell updates/sec
Tabular output not generated.
```

Scoring table: PAM 150  
Gap 11

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Statistics: Mean 44.623; Variance 88.995; scale 0.501

## SUMMARY

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	364	24.3	204	2	S60325	transcription factor	1.01e-46
2	276	18.5	304	2	D72340	probable zinc finger	4.13e-30
3	263	17.6	180	2	D71448	probable zn finger pr	1.01e-27
4	169	11.3	209	2	S55887	CCHH finger protein 7	2.60e-11
5	164	11.0	228	2	S55881	CCHH finger protein 1	1.69e-10
6	160	10.7	197	2	S55886	CCHH finger protein 6	7.49e-10
7	157	10.5	259	2	S55884	CCHH finger protein 4	2.26e-09
8	154	10.3	150	2	S55882	CCHH finger protein 2	6.80e-09
9	148	9.9	235	2	S55883	CCHH finger protein 3	5.98e-08
10	141	9.4	228	2	S55888	CCHH finger protein 8	7.23e-07
11	125	8.4	211	2	S55885	CCHH finger protein 5	1.77e-04
12	119	8.0	273	2	D19885	zinc-finger protein,	1.29e-03
13	116	7.8	281	2	S319159	finger protein EPF1 -	3.40e-03
14	117	7.8	485	2	A40751	finger protein MZF1 -	2.46e-03
15	115	7.7	339	2	JC1442	transcription factor	4.69e-03
16	113	7.6	339	2	B34895	transcription factor	8.87e-03
17	114	7.6	354	2	D148722	zinc finger protein -	6.45e-03
18	113	7.6	365	2	A34895	5S RNA-binding protei	8.87e-03
19	114	7.6	455	2	D148724	zinc finger protein p	6.45e-03
20	113	7.6	594	2	JC5146	arylphorin gene-speci	8.87e-03
21	112	7.5	273	2	S69193	probable finger prote	1.22e-02
22	110	7.4	308	2	A44496	transcription repress	2.28e-02
23	111	7.4	484	2	I3920020	zinc finger protein C	1.67e-02

ORGANISM	#formal_name Arabidopsis thaliana #common_name mouse-ear cress			
DATE	05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Mar-1999			
ACCESSIONS	T02540			
REFERENCE	Z14198			
#authors	Rounisley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.			
#submission	submitted to the EMBL Data Library, June 1998			
#description	Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.			
#accession	T02540			
##status	preliminary; translated from GB/EMBL/DBJ			
##molecule_type	DNA			
##residues	1-304 #label ROU			
##cross-references	EMBL:AC004684; NID:g3236234; PID:g3236256			
GENETICS				
#map_position	2			
#note	F13M22.24			
SUMMARY	#length 304 #molecular-weight 34125 #checksum 1987			
Query Match	18.5%; Score 276; DB 2; Length 304;			
Best Local Similarity	86.8%; Pred. No. 4.13e-30;			
Matches	33; Conservative 4; Mismatches 1; Indels 0; Gaps 0;			
Db	34 WPPRSYTCFCRRFSAQALGGHNVHRDRARLQKA 71			
Qy	61 WPPRSYTCFCRRFSAQALGGHNVHRDRARLQKS 98			
RESULT	3			
ENTRY	D71448 #type complete			
TITLE	probable zn finger protein - Arabidopsis thaliana			
ORGANISM	#formal_name Arabidopsis thaliana #common_name mouse-ear cress			
#variety	columbia			
DATE	03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998			
ACCESSIONS	D71448			
REFERENCE	A71400			
#authors	Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.; Stiekema, L.; Drost, R.; Ridley, P.; Hudson, S.A.; Patel, K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.; Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka, A.; Vouklatou, E.; Milioni, D.; Hatzopoulos, P.; Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reckman, S.; Ansgorge, W.; Cooke, R.; Berger, C.; Delseny, M.; Voet, M.; Volckaert, G.; Meves, H.W.; Klosterman, S.; Schueller, C.; Chalwatzis, N.			
#journal	Nature (1998) 391:485-488			
#title	Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana.			
#cross-references	MUID:98121113			
#accession	D71448			
##status	preliminary; nucleic acid sequence not shown; translation not shown			
##molecule_type	DNA			
##residues	1-180 #label BEV			
##cross-references	GB:297344; NID:g2245126; PID:e327088; PID:g2245140			
GENETICS				
#map_position	4COP9-4G3845			
SUMMARY	#length 180 #molecular-weight 19788 #checksum 748			
Query Match	17.6%; Score 263; DB 2; Length 180;			
Best Local Similarity	86.5%; Pred. No. 1.01e-27;			



[illegible]





THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Wed Jul 14 17:25:00 1999: Maspar time 8.03 seconds

```
Run on: Wed Jul 14 17:25:00 1999; MasPar time 8.03 Seconds
725.174 Million cell updates/sec
Tabular output not generated.
```

```
>US-09-156-580-2
Title:
Description: (1-206) from US09156580A.pcp
Perfect Score: 1495
Sequence: 1 MAGDRNDSFNISKYFKNKSIM.....IGLISSEKEDDLLELRGST 206
```

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

```
Database: swiss-prot37
          1:swissprot
```

Statistics: Mean 45.702; Variance 77.977; scale 0.586

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	117	7.8	485	1	ZNF42_HUMAN	ZINC FINGER PROTEIN 42	2.27e-04	
2	115	7.7	339	1	TF3A2_BUFAM	TRANSCRIPTION FACTOR 1	4.74e-04	
3	115	7.7	397	1	YKQ8_CAEEL	HYPOTHETICAL ZINC FING	4.74e-04	
4	113	7.6	365	1	P43_XENBO	P43 5S RNA BINDING PRO	9.84e-04	
5	110	7.4	308	1	A6F1_DROME	ADULT ENHANCER FACTOR	2.90e-03	
6	111	7.4	512	1	Z1169_HUMAN	ZINC FINGER PROTEIN 16	2.03e-03	
7	110	7.4	3005	1	ZFH2_DROME	ZINC-FINGER PROTEIN 2	2.90e-03	
8	109	7.3	1350	1	XFIN_XENLA	XFIN PROTEIN.	4.14e-03	
9	107	7.2	289	1	ZN75_HUMAN	ZINC FINGER PROTEIN 75	8.39e-03	
10	108	7.2	704	1	BC3_DROME	BROAD-COMPLEX CORE-NS-	5.90e-03	
11	108	7.2	899	1	SUHW_DROVI	SUPPRESSOR OF HAIRY WI	5.90e-03	
12	105	7.0	196	1	Z68_XENLA	GASTRULA ZINC FINGER P	1.69e-02	
13	104	7.0	264	1	SNAI_MOUSE	SNAIL PROTEIN HOMOLOG	2.38e-02	
14	105	7.0	392	1	ODL_DROME	ODD-SKIPPED PROTEIN.	1.69e-02	
15	104	7.0	423	1	ZN1_RAT	ZINC FINGER PROTEIN GF	2.38e-02	
16	104	7.0	630	1	MN4_YEAST	ZINC FINGER PROTEIN MS	2.38e-02	
17	104	7.0	1355	1	SALM_DROME	HOMEOTIC PROTEIN SPALT	2.38e-02	
18	104	7.0	1402	1	SALM_DROVI	HOMEOTIC PROTEIN SPALT	2.38e-02	
19	103	6.9	1191	1	ZN91_HUMAN	ZINC FINGER PROTEIN 91	3.36e-02	
20	101	6.8	604	1	GLAS_DROME	GLASS PROTEIN.	6.63e-02	
21	102	6.8	625	1	R101_YEAST	REGULATORY PROTEIN RIM	4.73e-02	
22	101	6.8	636	1	ZF90_MOUSE	ZINC FINGER PROTEIN 90	6.63e-02	
23	102	6.8	728	1	Y413_MYCGE	HYPOTHETICAL PROTEIN M	4.73e-02	

[illegible]

[illegible]

```
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97225201.
RA CHIDAMBARAM A., GAILANI M., GERRARD B., STEWART C., GOLDSTEIN A.,
RA CHUMAKOV I., BALE A.E., DEAN M.;
RT "Characterization of a YAC contig containing the NBCCS locus and a
RT novel Kruppel-type zinc finger sequence on chromosome segment
RT 9q22.3";
RL GENES CHROMOSOMES CANCER 18:212-218(1997).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN KIDNEY, IS WEAKLY
CC EXPRESSED IN HEART, LIVER, SPLEEN, AND SMALL INTESTINE, AND IS NOT
CC EXPRESSED IN ADULT BRAIN OR SPINAL CORD.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28251; G903599; -.
CC DR MM; 603404; -.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2; 9.
CC DR PFAM; PF00096; zf-C2H2; 11.
CC KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; NUCLEAR PROTEIN.
CC FT NON_TER 1 1
CC FT DOMAIN 144 473 ZINC-FINGERS.
CC FT ZN_FING 144 166 C2H2-TYPE.
CC FT ZN_FING 172 194 C2H2-TYPE.
CC FT ZN_FING 200 222 C2H2-TYPE.
CC FT ZN_FING 228 250 C2H2-TYPE.
CC FT ZN_FING 311 333 C2H2-TYPE.
CC FT ZN_FING 339 361 C2H2-TYPE.
CC FT ZN_FING 367 389 C2H2-TYPE.
CC FT ZN_FING 395 417 C2H2-TYPE.
CC FT ZN_FING 449 473 C2H2-TYPE.
CC FT ZN_FING 479 504 C2H2-TYPE.
CC SQ SEQUENCE 512 AA; 57651 MW; 0FA6FE76 CRC32;
CC -----
Query Match 7.48; Score 111; DB 1; Length 512;
Best Local Similarity 29.88; Pred.No. 2.03e-03;
Matches 14; Conservative 13; Mismatches 19; Indels 1; Gaps 1;
Db 198 KPVYRCGRRHFRYTSLLTNHKKRIHSGRPFVQCQCGRFRQKIAL 244
: : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 64 RSYTCFCKREFRSAQALGCHMVNRDRILRQSPPRDI-NRYSLL 109
: : | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 7
ID ZFH2_DROME STANDARD; PRT; 3005 AA.
AC P28167;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN ZFH-2.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92001539.
RA FORTINI M.E., LAI Z., RUBIN G.M.;
RT "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
RT containing both zinc-finger and homeodomain motifs.";
RL MECH. DEV. 34:113-122(1991).
```

```
CC -!- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC NERVOUS SYSTEM.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.
CC -!- SIMILARITY: CONTAINS THREE HOMEBOX DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63450; G158823; -.
CC DR PIR; S27817; S27817.
CC DR PIR; S33642; S33642.
CC DR FLYBASE; FBgn0004607; zfh2.
CC DR PROSITE; PS00027; HOMEBOX_1; 2.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
CC DR PROSITE; PS00071; HOMEBOX_2; 3.
CC DR PFAM; PF00046; homeobox; 3.
CC DR PFAM; PF00096; zf-C2H2; 12.
CC DR HSSP; P15822; 4ZNF.
CC DR TRANSFAC; T00920; -.
CC KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; HOMEBOX; NUCLEAR PROTEIN;
CC REPEAT.
CC FT ZN_FING 133 156 C2H2-TYPE.
CC FT ZN_FING 559 582 C2H2-TYPE.
CC FT ZN_FING 614 638 C2H2-TYPE.
CC FT ZN_FING 732 756 C2H2-TYPE.
CC FT ZN_FING 897 916 C2H2-TYPE (DEGENERATE).
CC FT ZN_FING 940 964 C2H2-TYPE.
CC FT ZN_FING 999 1023 C2H2-TYPE.
CC FT ZN_FING 1074 1098 C2H2-TYPE.
CC FT ZN_FING 1210 1233 C2H2-TYPE.
CC FT ZN_FING 1341 1365 C2H2-TYPE.
CC FT ZN_FING 1438 1462 C2H2-TYPE.
CC FT ZN_FING 1477 1500 C2H2-TYPE (DEGENERATE).
CC FT ZN_FING 1513 1535 C2H2-TYPE.
CC FT ZN_FING 1541 1564 C2H2-TYPE.
CC FT DNA_BIND 1797 1856 HOMEBOX 1.
CC FT DNA_BIND 2154 2213 HOMEBOX 2.
CC FT ZN_FING 2234 2256 C2H2-TYPE.
CC FT ZN_FING 2371 2393 C2H2-TYPE.
CC FT DNA_BIND 2760 2819 HOMEBOX 3.
CC SQ SEQUENCE 3005 AA; 332056 MW; 8B4CC45F CRC32;
CC -----
Query Match 7.48; Score 110; DB 1; Length 3005;
Best Local Similarity 68.48; Pred.No. 2.90e-03;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db 1543 CSFCQNFNRSTQALQKME 1561
||||:|||||:|:|
QY 68 CSFCREFRSAQALGGHMN 86
RESULT 8
ID XFIN_XENLA STANDARD; PRT; 1350 AA.
AC P08045;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE XFIN PROTEIN.
GN XFIN.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-EMBRIO;
RX MEDLINE; 88082679.
RA RUIZ I ALTABA R., PERRY-O'KEEFE H., MELTON D.A.;
```



RT "xfin: an embryonic gene encoding a multifingered protein in  
RT Xenopus";  
RL EMBO J. 6:3065-3070(1987).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE: 94021366.  
RA ANDREAZZOLI M., DE LUCCHINI S., COSTA M., BARSACCHI G.;  
RT "RNA binding properties and evolutionary conservation of the xenopus  
RT multifinger protein Xfin";  
RL NUCLEIC ACIDS RES. 21:4218-4225(1993).  
RN [3]  
RP STRUCTURE BY NMR OF FINGER 31.  
RX MEDLINE: 89346749.  
RA LEE M.S., GIPPERT G.P., SOMAN K.V., CASE D.A., WRIGHT P.E.;  
RT "Three-dimensional solution structure of a single zinc finger DNA-  
RT binding domain";  
RL SCIENCE 245:635-637(1989).  
RN [4]  
RP STRUCTURE BY NMR OF A FINGER.  
RX MEDLINE: 89378224.  
RA LEE M.S., CAVANAGH J., WRIGHT P.E.;  
RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc  
RT finger from Xfin. Sequential resonance assignments and secondary  
RT structure";  
RL FEBS LETT. 254:159-164(1989).  
RN [5]  
RP FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL  
CC REGULATION PROCESSES.  
CC -!- SUBCELLULAR LOCATION: CYTOSOL.  
CC -!- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH  
CC AS NEURAL RETINA CONES.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND  
CC EMBRYOGENESIS.  
CC -!- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.  
CC -!- PTM: PHOSPHORYLATED.  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: CONTAINS 1 KRAB BOX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X06021; G55234; -  
CC PIR: S00647; S00647.  
CC PDB: 1ZNF; 15-OCT-91.  
CC PROSITE: PS00028; ZINC\_FINGER\_C2H2; 35.  
CC PFAM: PF00096; zf-C2H2; 36.  
KW ZINC-FINGER; METAL-BINDING; RNA-BINDING; REPEAT; 3D-STRUCTURE;  
KW PHOSPHORYLATION.  
FT DOMAIN 1 26 KRAB BOX ("A BOX").  
FT DOMAIN 27 58 KRAB BOX ("B BOX").  
FT DOMAIN 108 298 ZINC-FINGERS I.  
FT DOMAIN 326 488 ZINC-FINGERS II.  
FT DOMAIN 503 721 ZINC-FINGERS III.  
FT DOMAIN 750 940 ZINC-FINGERS IV.  
FT DOMAIN 988 1066 ZINC-FINGERS V.  
FT DOMAIN 1136 1298 ZINC-FINGERS VI.  
FT ZN\_FING 108 130 C2H2-TYPE.  
FT ZN\_FING 136 158 C2H2-TYPE.  
FT ZN\_FING 164 186 C2H2-TYPE.  
FT ZN\_FING 192 214 C2H2-TYPE.  
FT ZN\_FING 220 242 C2H2-TYPE.  
FT ZN\_FING 248 270 C2H2-TYPE.  
FT ZN\_FING 276 298 C2H2-TYPE.  
FT ZN\_FING 326 348 C2H2-TYPE.  
FT ZN\_FING 354 376 C2H2-TYPE.  
FT ZN\_FING 382 404 C2H2-TYPE.  
FT ZN\_FING 410 432 C2H2-TYPE.  
FT ZN\_FING 438 460 C2H2-TYPE.

FT ZN\_FING 466 488 C2H2-TYPE.  
FT ZN\_FING 503 525 C2H2-TYPE.  
FT ZN\_FING 531 553 C2H2-TYPE.  
FT ZN\_FING 559 581 C2H2-TYPE.  
FT ZN\_FING 587 609 C2H2-TYPE.  
FT ZN\_FING 615 637 C2H2-TYPE.  
FT ZN\_FING 643 665 C2H2-TYPE.  
FT ZN\_FING 671 693 C2H2-TYPE.  
FT ZN\_FING 699 721 C2H2-TYPE.  
FT ZN\_FING 750 772 C2H2-TYPE.  
FT ZN\_FING 778 800 C2H2-TYPE.  
FT ZN\_FING 806 828 C2H2-TYPE.  
FT ZN\_FING 834 856 C2H2-TYPE.  
FT ZN\_FING 862 884 C2H2-TYPE.  
FT ZN\_FING 890 912 C2H2-TYPE.  
FT ZN\_FING 918 940 C2H2-TYPE.  
FT ZN\_FING 988 1010 C2H2-TYPE.  
FT ZN\_FING 1016 1038 C2H2-TYPE.  
FT ZN\_FING 1044 1066 C2H2-TYPE.  
FT ZN\_FING 1136 1158 C2H2-TYPE.  
FT ZN\_FING 1164 1186 C2H2-TYPE.  
FT ZN\_FING 1192 1214 C2H2-TYPE.  
FT ZN\_FING 1220 1242 C2H2-TYPE.  
FT ZN\_FING 1248 1270 C2H2-TYPE.  
FT ZN\_FING 1276 1298 C2H2-TYPE.  
FT STRAND 1045 1045  
FT STRAND 1052 1052  
FT HELIX 1056 1065  
FT TURN 1066 1066  
SQ SEQUENCE 1350 AA; 155804 MW; 10CA7C8C CRC32;  
Query Match 7.3%; Score 109; DB 1; Length 1350;  
Best Local Similarity 37.5%; Pred. No. 4.14e-03;  
Matches 12; Conservative 8; Mismatches 12; Indels 0; Gaps 0;  
Db 1042 RPYKGLGCRSEVKSALSRHQVHKNEPVL 1073  
QY 64 RSYTCSFCKREFRSQAUGGHMVRDRRAIL 95  
RESULT 9  
ID ZN75\_HUMAN STANDARD; PRT; 289 AA.  
AC P51815;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE ZINC FINGER PROTEIN 75.  
GN ZNF75.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG FIBROBLAST;  
RX MEDLINE: 94116987;  
RA VILLA A., ZUCCHI I., PILIA G., STRINA D., SUSANI L., MORALI F.,  
RA PATROSSO C., FRATTINI A., LUCCHINI F., REPETTO M., SACCO M.G.,  
RA ZOPPE M., VEZZONI P.;  
RT "ZNF75: isolation of a cDNA clone of the KRAB zinc finger gene  
RT subfamily mapped in YACs 1 Mb telomeric of HPRT";  
RL GENOMICS 18:223-229(1993).  
RN [2]  
RP SEQUENCE OF 139-289 FROM N.A.  
RA MARINO M., ARCHIDIAONO N., FRANZE N., ROSATI M., ROCCHI M.,  
RA BALLABIO A., GRIMALDI G.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: CONTAINS 1 KRAB BOX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; S67970; G460903; -  
DR EMBL; X68010; G525244; -  
DR MIM; 314997; -  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 5.  
DR PFAM; PF00096; zf-C2H2; 5.  
DR HSP; P25490; IZNM.  
KW TRANSCRIPTION REGULATION; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.  
FT DOMAIN 16 55 KRAB BOX.  
FT DOMAIN 144 278 ZINC-FINGERS.  
FT ZN\_FING 144 166 C2H2-TYPE.  
FT ZN\_FING 172 194 C2H2-TYPE.  
FT ZN\_FING 200 222 C2H2-TYPE.  
FT ZN\_FING 228 250 C2H2-TYPE.  
FT ZN\_FING 256 278 C2H2-TYPE.  
SQ SEQUENCE 289 AA; 33683 MW; 569284FB CRC32;

Query Match 7.2%; Score 107; DB 1; Length 289;  
Best Local Similarity 46.9%; Pred. No. 8.39e-03;  
Matches 15; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Db 254 QPYTCSLKRNFSRRSSLLRHOKLRRREACL 285

QY 64 RSYTCSFKRFSQAALGGHNVHRRDRAIL 95

RESULT 10  
ID BRC3\_DROME STANDARD; PRT; 704 AA.  
AC Q01293;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE BROAD-COMPLEX CORE-NS-23 PROTEIN (BRCORE-NS-23).  
GN BR OR BR-C.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OREGON-R.  
RX MEDLINE; 92077389.  
RA DIBELLO P.R., WITHERS D.A., BAYER C.A., FRISTROM J.W., GUILD G.M.;  
RT "The Drosophila Broad-Complex encodes a family of related proteins  
RT containing zinc fingers."  
RL GENETICS 129:385-397(1991).  
CC -!- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND  
CC TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR PRODUCTS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING, BRCORE-NS-23 (SHOWN HERE), BRCORE-TNT1-Q1-Z1  
CC (AC Q01295), BRCORE-Q1-Z1 (AC Q01296) AND BRCORE-Z2 (AC Q01296). A  
CC COMMON CORE N-TERMINAL DOMAIN IS USED TO ALTERNATE ZINC-FINGER C-  
CC TERMINAL DOMAINS.  
CC -!- SIMILARITY: CONTAINS A N-TERMINAL BR-C/TTK DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X54664; G10901; -  
DR FLYBASE; FBgn0000210; br.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 2.

DR PFAM; PF00096; zf-C2H2; 2.  
DR PFAM; PF00651; BTB; 1.  
DR TRANSFAC; T01479; -  
KW NUCLEAR PROTEIN; DNA-BINDING; DEVELOPMENTAL PROTEIN; ZINC-FINGER;  
KW METAL-BINDING; ALTERNATIVE SPLICING.  
FT DOMAIN 1 115 BR-C/TTK DOMAIN.  
FT DOMAIN 203 207 POLY-ALA.  
FT DOMAIN 265 268 POLY-ASN.  
FT DOMAIN 435 551 ASP/SER-RICH.  
FT DOMAIN 540 544 POLY-ASN.  
FT DOMAIN 545 551 POLY-SER.  
FT DOMAIN 587 640 ZINC-FINGERS.  
FT ZN\_FING 587 610 C2H2-TYPE.  
FT ZN\_FING 617 640 C2H2-TYPE.  
SQ SEQUENCE 704 AA; 74468 MW; 9CC18352 CRC32;

Query Match 7.2%; Score 108; DB 1; Length 704;

Best Local Similarity 30.3%; Pred. No. 5.90e-03;  
Matches 10; Conservative 12; Mismatches 10; Indels 1; Gaps 1;

Db 617 YVCFCHRRYRTKNSLTTHKSLQHRGSSGMLKR 649

QY 66 YTCSCFKRFSQAALGGHNVHRRDRA-ILRQ 97

RESULT 11  
ID SUHW\_DROVI STANDARD; PRT; 899 AA.  
AC Q08876;  
DT 01-OCT-1994 (REL. 30, CREATED)  
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE SUPPRESSOR OF HAIRY WING PROTEIN.  
GN SU(HW).  
OS DROSOPHILA VIRILIS (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94010293.  
RA HARRISON D.A., GDULA D.A., COYNE R.S., CORCES V.G.;  
RT "A leucine zipper domain of the suppressor of hairy-wing protein  
RT mediates its repressive effect on enhancer function."  
RL GENES DEV. 7:1966-1978(1993).  
CC -!- FUNCTION: SU(HW) CONTROLS THE PHENOTYPIC EFFECT OF THE GYPSY  
CC TRANSPOSABLE ELEMENT. BINDS SPECIFICALLY TO A REGION OF THE GYPSY  
CC ELEMENT LOCATED 3' OF THE 5'LTR. IT IS PROBABLY A TRANSCRIPTION  
CC FACTOR. COULD PLAY A ROLE IN THE ESTABLISHMENT OF CHROMATIN  
CC DOMAINS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z25520; G396677; -  
DR PIR; S36436; S36436.  
DR PIR; B48586; B48586.  
DR FLYBASE; FBgn0013144; Dvir\su(HW).  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 10.  
DR PFAM; PF00096; zf-C2H2; 12.  
DR HSP; P25490; I0BD.  
KW TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;  
KW REPEAT; NUCLEAR PROTEIN.  
FT DOMAIN 218 617 ZINC-FINGERS.  
FT ZN\_FING 218 240 C2H2-TYPE.  
FT ZN\_FING 288 311 C2H2-TYPE.  
FT ZN\_FING 318 340 C2H2-TYPE.  
FT ZN\_FING 347 365 C2H2-TYPE.

```

FT ZN_FING 379 401 C2H2-TYPE.
FT ZN_FING 412 434 C2H2-TYPE.
FT ZN_FING 440 462 C2H2-TYPE.
FT ZN_FING 468 490 C2H2-TYPE.
FT ZN_FING 496 518 C2H2-TYPE.
FT ZN_FING 522 544 C2H2-TYPE.
FT ZN_FING 552 576 C2H2-TYPE.
FT ZN_FING 594 617 C2H2-TYPE.
SQ SEQUENCE 899 AA; 100592 MW; A884EA7A CRC32;

Query Match 7.2%; Score 108; DB 1; Length 899;
Best Local Similarity 29.6%; Pred. No. 5.90e-03;
Matches 16; Conservative 16; Mismatches 20; Indels 2; Gaps 2;

Db 520 RPYACSLCQFKTEKQLERHVKHQTRKRASTACTECTRTSFRFSALLKEHLDA 573
Qy 64 RSYTCSCFKREFRSQAALGGHMNVH-RRDRAILL-RQSPPRDINRYSLNLEP 115

RESULT 12
ID ZGB XENLA STANDARD; PRT; 196 AA.
AC P18737;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GASTRULA ZINC FINGER PROTEIN XLCGF8.2DB (FRAGMENT).
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90040698.
RA NIETZEL W., EL-BARADI T., MENTZEL H., PIELER T., KOESTER M.,
RA POSTING A., KNOESCHEL W.;
RT "Second-order repeats in Xenopus laevis finger proteins.";
J. MOL. BIOL. 208:639-659(1989).
DR PIR: S06561; S06561.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 7.
DR PFAM: PF00096; zf-C2H2; 7.
DR HSP: P23490; 12NM.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT.
FT NON_TER 1 1
FT ZN_FING 6 28 C2H2-TYPE.
FT ZN_FING 34 56 C2H2-TYPE.
FT ZN_FING 62 84 C2H2-TYPE.
FT ZN_FING 90 112 C2H2-TYPE.
FT ZN_FING 118 140 C2H2-TYPE.
FT ZN_FING 146 168 C2H2-TYPE.
FT ZN_FING 174 196 C2H2-TYPE.
FT NON_TER 196 196
SQ SEQUENCE 196 AA; 22023 MW; 3799487C CRC32;

Query Match 7.0%; Score 105; DB 1; Length 196;
Best Local Similarity 24.3%; Pred. No. 1.69e-02;
Matches 17; Conservative 25; Mismatches 25; Indels 3; Gaps 3;

Db 98 EFAKHRLGLHKLHITGKPSCTCGHFAKHVHLVSHMKIHTREKPTCTCEGEHAN 157
Qy 47 NYGNEDHLGLL-FSWPPRSYTCSCFKREFRSQAALGGHMNVHRRDRAILL-RQSPPRDIN 104
Db 158 KVSLLGLHLM 167
Qy 105 RYSLN-LNLEP 113

RESULT 13
ID SNAIL_MOUSE STANDARD; PRT; 264 AA.
AC Q02085;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SNAIL PROTEIN HOMOLOG (SNA PROTEIN).
GN SNA.

```

```

OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-EMBRYO;
RX MEDLINE: 93201990.
RA SMITH D.E., DEL AMO F.F., GRIDLEY T.;
RT "Isolation of Sna, a mouse gene homologous to the Drosophila genes
RT snail and escargot: its expression pattern suggests multiple roles
RT during postimplantation development.";
RL DEVELOPMENT 116:1033-1039(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93130772.
RA NIETO A.M., BENNETT M.F., SARGENT M.G., WILKINSON D.G.;
RT "Cloning and developmental expression of Sna, a murine homologue of
RT the Drosophila snail gene.";
RL DEVELOPMENT 116:227-237(1992).
RN [3]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN-129/SV;
RA JIANG R., COPELAND N.G., GILBERT D.J., JENKINS N.A., GRIDLEY T.;
RA SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: THIS PROTEIN HAS MANY ROLES DURING POSTIMPLANTATION
CC DEVELOPMENT. IT IS INVOLVED IN EMBRYONIC MESODERM FORMATION
CC AND ITS MAINTENANCE AND MAY ALSO BE INVOLVED IN CHONDROGENESIS
CC AND IN EPITHELIAL-MESENCHYMAL INDUCTIVE INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: POSTIMPLANTATION.
CC -1- SIMILARITY: TO D.MELANOGASTER SNAIL PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M95604; G429188;
CC DR EMBL: X67253; G54121;
CC DR EMBL: U95961; G2105425;
CC DR MGD: MGI:98330; SNA.
CC DR PROSITE: PS00028; ZINC_FINGER_C2H2; 3.
CC DR PFAM: PF00096; zf-C2H2; 4.
CC DR HSP: P08047; 1SP2.
CC DR TRANSFAC: T02333;
CC DEVELOPMENTAL PROTEIN: ZINC-FINGER; METAL-BINDING; DNA-BINDING;
KW NUCLEAR PROTEIN; REPEAT.
FT DOMAIN 236 259 ZINC-FINGERS.
FT ZN_FING 156 176 C2H2-TYPE.
FT ZN_FING 180 202 C2H2-TYPE.
FT ZN_FING 210 230 C2H2-TYPE.
FT ZN_FING 236 259 C2H2-TYPE.
FT CONFLICT 219 219 D -> V (IN REF. 2).
SQ SEQUENCE 264 AA; 29190 MW; 6926770 CRC32;

Query Match 7.0%; Score 104; DB 1; Length 264;
Best Local Similarity 28.8%; Pred. No. 2.38e-02;
Matches 15; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

Db 206 KPFSCHCNRAFADRSNLRALHQTSDVKRYQCACARTFSRMSLLHKHQS 257
Qy 64 RSYTCSCFKREFRSQAALGGHMNVHRRDRAILL-RQSPPRDINRYSLNLEP 115

RESULT 14
ID ODD_DROME STANDARD; PRT; 392 AA.
AC P23803;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

```





```

RESULT 2
ID O80942 PRELIMINARY; PRT; 304 AA.
AC O80942;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE ZINC FINGER PROTEIN.
GN F13M22.24.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA BRANSON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.";
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC004684; G3236256; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 304 AA; 34125 MW; 879B713E CRC32;

Query Match 19.5%; Score 276; DB 10; Length 304;
Best Local Similarity 86.8%; Pred. No. 1.01e-34;
Matches 33; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 34 WPPRSYTCSCRRCKFSQAALGGHMVHRRDRARLQKA 71
Qy 61 WPPRSYTCSCRRCKFSQAALGGHMVHRRDRARLQKS 98

RESULT 3
ID O23621 PRELIMINARY; PRT; 180 AA.
AC O23621;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 2N FINGER PROTEIN HOMOLOG.
DE ZN FINGER PROTEIN HOMOLOG.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE; 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERGKAMP R., DIRKSE W., VAN STAVEREN M., STICKEN W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFEANELLI P., WEDLER H.,
RA WEDLER E., WAMBUITT R., WEITZENGGER T., FOHL T.M., TERRY N.,
RA GLESEN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOUKELATOU E., MILIONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA PALME K., BENES V., RECHMAN S., ANSGORGE W., COOKE R., BERGER C.,
RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RA Arabidopsis thaliana."
RL NATURE 391:485-488(1998).
DR EMBL; 297344; E327088; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.

```



```
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LANDSBERG RECTA; TISSUE=ROOT;
RX MEDLINE; 95322589.
RA TAGUE B.W., GOODMAN H.M.;
RT "Characterization of a family of Arabidopsis zinc finger protein
  CDNAS.";
RL PLANT MOL. BIOL. 28:267-279(1995).
DR EMBL; L39646; G790677; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR PFAM; PF00096; zf-C2H2; 1.
DR MENDEL; 16156; ARATH;1523;mn16156.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 235 AA; 26351 MW; BEEC55BF CRC32;

Query Match          9.4%; Score 148; DB 10; Length 235;
Best Local Similarity 51.7%; Pred. No. 1.96e-09;
Matches 15; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

Db 61 FSCNYCQRTFYSSQALGGHQHNAKRRTLAKRG 93
      :|:|:| | | | | | | | | | | | | | | | |
Qy 66 YTCSCFKREFRSQAALGGHMVHRRDRAILROS 98

RESULT 10
ID Q39267 PRELIMINARY; PRT; 239 AA.
AC Q39267;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN (FRAGMENT).
GN ZFP8.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LANDSBERG RECTA; TISSUE=LEAF;
RX MEDLINE; 95322589.
RA TAGUE B.W., GOODMAN H.M.;
RT "Characterization of a family of Arabidopsis zinc finger protein
  CDNAS.";
RL PLANT MOL. BIOL. 28:267-279(1995).
DR EMBL; L39651; G790687; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
FT NON_TER 1
SQ SEQUENCE 239 AA; 27781 MW; 48976A9F CRC32;

Query Match          9.4%; Score 141; DB 10; Length 239;
Best Local Similarity 51.7%; Pred. No. 3.37e-08;
Matches 15; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 91 RRFECYCFRNFTPSQALGGHQHNAKRR 119
      | : | : | : | : | : | : | : | : | : |
Qy 64 RSYTCSCFKREFRSQAALGGHMVHRRDR 92

RESULT 11
ID P93751 PRELIMINARY; PRT; 257 AA.
AC P93751;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN 8, ZFP8.
GN T06D20.16.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPALES; BRASSICACEAE; ARABIDOPSIS.
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., PHILLIPS C.A., BRANDON R.C.,
RA FUHRMANN J.L., WHITE O., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA VENTER J.C.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U90439; G1871188; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 257 AA; 29571 MW; FB769546 CRC32;

Query Match          9.4%; Score 141; DB 10; Length 257;
Best Local Similarity 51.7%; Pred. No. 3.37e-08;
Matches 15; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 93 RRFECYCFRNFTPSQALGGHQHNAKRR 121
      | : | : | : | : | : | : | : | : | : |
Qy 64 RSYTCSCFKREFRSQAALGGHMVHRRDR 92

RESULT 12
ID Q22086 PRELIMINARY; PRT; 166 AA.
AC Q22086;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ZPT2-14.
OS PETUNIA X HYBRIDA.
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; PETUNIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MITCHELL DIPLOID; TISSUE=ROOT;
RX MEDLINE; 98083196.
RA KUBO K., SAKAMOTO A., KOBAYASHI A., RYBKA Z., KANNO Y., NAKAGAWA H.,
RA NISHINO T., TAKATSUJI H.;
RT "Cys2/His2 zinc-finger protein family of petunia: evolution and
  general mechanism of target-sequence recognition.";
RL NUCLEIC ACIDS RES. 26:608-615(1998).
DR EMBL; AB006601; D1022782; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
DR PFAM; PF00096; zf-C2H2; 2.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 166 AA; 18922 MW; 2EF440AC CRC32;

Query Match          8.8%; Score 132; DB 10; Length 166;
Best Local Similarity 32.6%; Pred. No. 1.20e-06;
Matches 15; Conservative 15; Mismatches 15; Indels 1; Gaps 1;

Db 27 FSHOENHFMNDSRPSRVECKTCNRQSFQALGGHRAHKKPR 72
      : | : | : | : | : | : | : | : | : |
Qy 48 YGNEDHLGLGF-SWPPRSYTCSCFKREFRSQAALGGHMVHRRDR 92

RESULT 13
ID P93714 PRELIMINARY; PRT; 341 AA.
AC P93714;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE PETHV, ZPT2-6.
OS PETUNIA X HYBRIDA.
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; PETUNIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STAMEN; STRAIN=CV. MITCHELL DIPLOID;
RA KOBAYASHI A., SAKAMOTO A., KUBO K., RYBKA Z., KANNO Y., TAKATSUJI H.;
RT "Seven zinc-finger transcription factors are expressed sequentially
```



RT during the development of anthers and pollen in Petunia.\*;  
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AB000452; D1019855; -;  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 2.  
DR PFAM; PF00096; zf-C2H2; 2.  
DR MENDEL; 12063; PETX; 1523; 2.  
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.  
SQ SEQUENCE 341 AA; 38431 MW; F10F3662 CRC32;

Query Match 8.6%; Score 129; DB 10; Length 341;  
Best Local Similarity 60.7%; Pred. No. 3.86e-06;  
Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Db 236 SPRIECSYCGAEFTSGOALGGHMRHR 263

QY 62 PPRSTCSFCKREFRSQAALGGHMMVHR 89

## RESULT 14

ID P93718 PRELIMINARY; PRT; 554 AA.  
AC P93718;  
DT 01-MAY-1997 (TRENBLREL. 03, CREATED)  
DT 01-JUL-1997 (TRENBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TRENBLREL. 09, LAST ANNOTATION UPDATE)  
DE PETHY, ZPT4-2.  
OS PETUNIA X HYBRIDA.  
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;  
OC ASTERIDAE; SOLANACEAE; SOLANALES; SOLANACEAE; PETUNIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-STAMEN; STRAIN-CV. MITCHELL DIPLOID;  
RA KOBAYASHI A., SAKAMOTO A., KUBO K., RYBKA Z., KANNO Y., TAKATSUI H.;  
RT Seven zinc-finger transcription factors are expressed sequentially during the development of anthers and pollen in Petunia.\*;  
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AB000456; D1020715; -;  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 3.  
DR PFAM; PF00096; zf-C2H2; 3.  
DR MENDEL; 12067; PETX; 1523; 6.  
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.  
SQ SEQUENCE 554 AA; 62932 MW; 8C0AA70D CRC32;

Query Match 8.6%; Score 129; DB 10; Length 554;  
Best Local Similarity 38.2%; Pred. No. 3.86e-06;  
Matches 21; Conservative 15; Mismatches 17; Indels 2; Gaps 2;

Db 488 KGHACPFGRMFKSGOALGGHGRSHFVSSNNHYQASAVOGKIVDLDLNL-PAP 541

QY 64 RSYTCSFCKREFRSQAALGGHMMVHRDRAILR-QSPPRDINRYSLLNLEPNP 117

## RESULT 15

ID P93719 PRELIMINARY; PRT; 554 AA.  
AC P93719; 004749;  
DT 01-MAY-1997 (TRENBLREL. 03, CREATED)  
DT 01-MAY-1997 (TRENBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-JAN-1999 (TRENBLREL. 09, LAST ANNOTATION UPDATE)  
DE TRANSCRIPTION FACTOR ZPT4-3.  
OS PETUNIA X HYBRIDA (GARDEN PETUNIA).  
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;  
OC ASTERIDAE; SOLANACEAE; SOLANALES; SOLANACEAE; PETUNIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. MITCHELL DIPLOID; TISSUE-STAMEN;  
RA KOBAYASHI A., SAKAMOTO A., KUBO K., RYBKA Z., KANNO Y., TAKATSUI H.;  
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; AB003672; D1020953; -;  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2; 4.  
DR PFAM; PF00096; zf-C2H2; 3.  
DR MENDEL; 12068; PETX; 1523; 7.

KW TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;  
KW NUCLEAR PROTEIN.  
FT ZN\_FING 9 31 C2H2-TYPE.  
FT ZN\_FING 119 141 C2H2-TYPE.  
FT ZN\_FING 417 439 C2H2-TYPE.  
FT ZN\_FING 488 510 C2H2-TYPE.  
SQ SEQUENCE 554 AA; 62616 MW; E7072B50 CRC32;

Query Match 8.6%; Score 128; DB 10; Length 554;

Best Local Similarity 39.3%; Pred. No. 5.67e-06;  
Matches 22; Conservative 13; Mismatches 18; Indels 3; Gaps 2;

Db 486 KGHECPYCDRVFKSGOALGGHGRSHFVSSNNHYQASAAKKEADDLDDLNL-PAP 540

QY 64 RSYTCSFCKREFRSQAALGGHMMVHRDRAI--LRQSPRDINRYSLLNLEPNP 117

Search completed: Wed Jul 14 17:25:46 1999  
Job time : 19 secs.

THIS PAGE BLANK (USPTO)